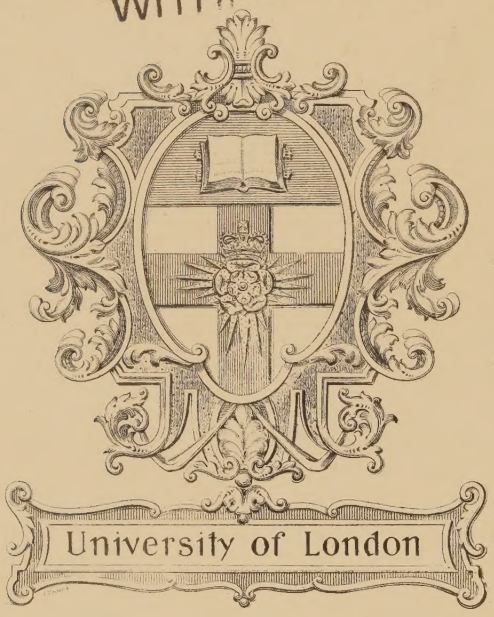
R PRIVATE
DEPOSITORY

SHL
WITHDRAWN



PROCEEDINGS

OF

THE LONDON MATHEMATICAL SOCIETY

SECOND SERIES

VOLUME 48

LONDON:
PRINTED AND PUBLISHED FOR THE SOCIETY BY
C. F. HODGSON & SON, LTD.,
2 NEWTON STREET, KINGSWAY, W.C.2.

1945

BIBL.
LONDIN.
UNIV.

CONTENTS

PAGE

Allen, H. S. Projective convergence and limit in sequence spaces ... 310
Cooper, J. L. B. The uniqueness of trigonometrical integrals ... 292
Cox, H. L., and Smith, H. E. The buckling of grids of stringers and ribs 1
————. The buckling of a thin sheet transversely stiffened 27
Edge, W. L. The contact net of quadrics 112
Fuchs, W. H. J. A uniqueness theorem for mean values of analytic functions 35
Goodstein, R. L. Function theory in an axiom-free equation calculus 401
Mahler, K. A problem of Diophantine approximation in quaternions 435
Mordell, L. J. On numbers represented by binary cubic forms ... 198
————. On the geometry of numbers in some non-convex regions... 339
Offord, A. C. An inequality for sums of independent random variables 467
Rado, R. Note on combinatorial analysis 122
Richardson, A. R. Groupoids and their automorphisms 83
Rogosinski, W. On the coefficients of subordinate functions ... 48
Smith, H. E., and Cox, H. L. The buckling of grids of stringers and ribs 1
————. The buckling of a thin sheet transversely stiffened ... 27
Todd, J. A. A table of partitions 229
Turing, A. M. A method for the calculation of the zeta-function ... 180
Walker, A. G. Relativistic mechanics: (III) Mechanics in 4-space, and conservative fields of force 161
Watson, G. N. A problem in elementary analysis proposed by Mordell 391
Whitehead, J. H. C. On the groups $\pi_r(V_{n,m})$ and sphere-bundles ... 243

PAPERS

PUBLISHED IN THE

PROCEEDINGS OF THE LONDON MATHEMATICAL SOCIETY

THE BUCKLING OF GRIDS OF STRINGERS AND RIBS

By H. L. Cox *and* H. E. Smith.

[Received 1 June, 1938.—Read 16 June, 1938.—Revised 23 February, 1939.]

Synopsis.

The paper describes a theoretical analysis of the stability of a plane rectangular grid consisting of two sets of parallel members (stringers and ribs) intersecting at right angles, when loaded in compression parallel to one set of members (the stringers). Both sets are composed of similar members of uniform section uniformly spaced and the ends of all members are held in the original plane of the grid. The exact solution of this problem appropriate to certain edge conditions is obtained both by consideration of the equations of equilibrium and by the Rayleigh-Ritz energy method.

Introduction.

The particular problem for which a complete solution is found is that of the buckling of a plane rectangular grid consisting of $m-1$ similar stringers of uniform section, each $2B$ long, spaced $2D/m(=2d)$ apart, and cross-connected by $n-1$ similar ribs of uniform section, each $2D$ long, spaced $2B/n$ $(=2b)$ apart, when subjected to uniform compression in the direction of the stringers. The ends of both stringers and ribs are connected to edge members having half the torsional rigidities of the corresponding main members, and these edge members are held in the original plane.

A problem of this type may be attacked by either of two methods. In the first, or "exact", method a system of variables is employed which can represent any conceivable buckled form of the structure, while another system of variables represents the applied forces and the interactions between the parts of the structure. The elasticity equations connecting these two systems are then formed, and the problem is completed by the solution of these equations. In the second method, known as the Rayleigh-Ritz energy method, an arbitrary form for the buckled structure is specified, this form being consistent with the boundary conditions and being capable of modification by a system of arbitrary variable parameters. The total strain energy due to this specified deformation is calculated, and the condition that this shall be a minimum is obtained by differentiation with respect to the parameters. This method is often called the approximate method, because, in its usual applications, the forms specified for the deformation are incapable of representing the actual deformation exactly. As a result, the deformed shape is represented only approximately, and in cases of stability the resulting estimate of the buckling load is necessarily greater than the true value. On the other hand, if the form chosen to represent the state of deformation is indeed capable of representing the true state of deformation, the Rayleigh-Ritz method must lead to the exact solution. The present problem provides an instance of an exact solution obtained by this method.

Part I. *Solution by the exact method.*

1. *Notation.* The $m+1$ stringers are numbered in succession $0 \to R \to m$ and the $n+1$ ribs $0 \to S \to n$; in the m and n ranges r and s respectively are used as current variables. At each intersection we define six quantities, viz. (at the intersection of the r-th stringer with the s-th rib):

w_{rs} = deflection,

ϕ_{rs} = slope of axis of stringer, or twist of rib (positive if increasing deflection in direction of s increasing),

ψ_{rs} = slope of axis of rib, or twist of stringer (positive if increasing deflection in direction of r increasing),

f_{rs} = force of stringer on rib (positive in direction of deflection),

m_{rs} = torque in rib transferred to stringer as bending moment (positive if tending to increase ϕ),

t_{rs} = torque in stringer transferred to rib as bending moment (positive if tending to increase ψ).

In addition to these $6mn-2m-2n-2$ unknowns, there is the further set of $m-1$ end loads, denoted by P_r. We shall assume, however, that these loads are connected by $m-2$ relations (*e.g.* $P_r = P$), so that in effect they represent but one variable.

2. *Buckling of a stringer.* We will first consider the equilibrium of the r-th stringer under the **above** system of forces. For convenience the suffix r will be omitted. It is assumed that buckling of the stringer is restricted to the plane of the forces and moments, and the positive directions of the forces and moments are taken as shown in Fig. 1.

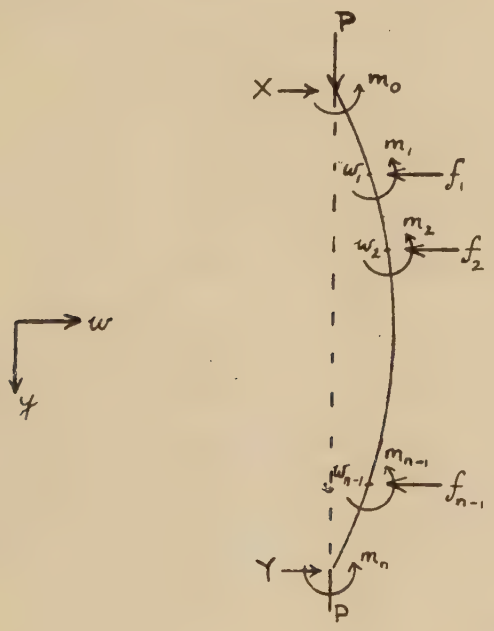


Fig. 1.

The lateral reaction X at $y = 0$ is then

$$\frac{1}{2B}\left\{\sum_0^n m_s + \sum_1^{n-1} f_s(y_n - y_s)\right\},$$

where y_s is the distance of the s-th rib from one end of the stringer. Then the deflection of the stringer over the range $y_{S-1} < y < y_S$ is determined by the equation

$$EI_S \frac{d^2w}{dy^2} = -Pw + Xy - m_0 - \ldots - m_{S-1} - f_1(y-y_1) - \ldots - f_{S-1}(y-y_{S-1}),$$

where I_S is the moment of inertia of the section of the strut in the range.

Restriction 1. Provided that I_S is constant in the range, this equation can be solved immediately, the solution being

$$Pw = -\sum_{0}^{S-1} m_s + \frac{y}{2B}\sum_{0}^{n} m_s + \frac{y}{2B}\sum_{1}^{n-1} f_s(y_n - y_s) - \sum_{1}^{S-1} f_s(y-y_s) + A_S \sin\sqrt{\left(\frac{P}{EI_S}\right)}\, y + B_S \cos\sqrt{\left(\frac{P}{EI_S}\right)}\, y.$$

Conditions of continuity of the stringer give $2(n-1)$ equations which may be used to eliminate the unknown coefficients A_S, B_S, etc.

Restriction 2. If we now assume that I_S is constant over the whole length of the stringer, and write

$$\sqrt{\left(\frac{P}{EI_S}\right)} \cdot \frac{2B}{n} = \theta,$$

hese equations of continuity become

$$(A_{S+1}-A_S)\sin(y_S n\theta/2B) + (B_{S+1}-B_S)\cos(y_S n\theta/2B) = m_S,$$

$$(A_{S+1}-A_S)\cos(y_S n\theta/2B) - (B_{S+1}-B_S)\sin(y_S n\theta/2B) = f_S \cdot 2B/(n\theta). \quad (1)$$

Hence

$$A_S = A_1 + \frac{2B}{n\theta}\sum_{1}^{S-1} f_s \cos\frac{y_s n\theta}{2B} + \sum_{0}^{S-1} m_s \sin\frac{y_s n\theta}{2B},$$

and

$$B_S = -\frac{2B}{n\theta}\sum_{1}^{S-1} f_s \sin\frac{y_s n\theta}{2B} + \sum_{0}^{S-1} m_s \cos\frac{y_s n\theta}{2B}. \quad (2)$$

Since $w = 0$ at $y = 0$, $B_1 = 0$; and since $w = 0$ at $y = 2B$,

$$A_n \sin n\theta + B_n \cos n\theta = 0.$$

Using this relation, we find

$$A_1 \sin n\theta = -\frac{2B}{n\theta}\sum_{1}^{n-1} f_s \sin\left(1-\frac{y_s}{2B}\right)n\theta - \sum_{0}^{n} m_s \cos\left(1-\frac{y_s}{2B}\right)n\theta; \quad (3)$$

so that, finally,

$$A_S = \frac{2B}{n\theta}\sum_{1}^{S-1} f_s \cos\frac{y_s n\theta}{2B} + \sum_{0}^{S-1} m_s \sin\frac{y_s n\theta}{2B} - \frac{1}{\sin n\theta}\left[\frac{2B}{n\theta}\sum_{1}^{n-1} f_s \sin\left(1-\frac{y_s}{2B}\right)n\theta + \sum_{0}^{n} m_s \cos\left(1-\frac{y_s}{2B}\right)n\theta\right], \quad (4a)$$

and

$$B_S = -\frac{2B}{n\theta}\sum_{1}^{S-1} f_s \sin\frac{y_s n\theta}{2B} + \sum_{0}^{S-1} m_s \cos\frac{y_s n\theta}{2B}. \quad (4b)$$

Using equations (4), we obtain

$$\frac{Pw_S}{2B} = \sum_1^{S-1} \left\{ \frac{y_s}{2B}\left(1-\frac{y_S}{2B}\right) - \frac{\sin(y_s n\theta/2B)\sin(1-y_S/2B)n\theta}{n\theta \sin n\theta} \right\} f_s$$
$$+ \sum_S^{n-1} \left\{ \frac{y_S}{2B}\left(1-\frac{y_s}{2B}\right) - \frac{\sin(y_S n\theta/2B)\sin(1-y_s/2B)n\theta}{n\theta \sin n\theta} \right\} f_s$$
$$- \sum_0^{S-1} \left\{ \left(1-\frac{y_S}{2B}\right) - \frac{\sin(1-y_S/2B)n\theta \cos(y_s n\theta/2B)}{\sin n\theta} \right\} \frac{m_s}{2B}$$
$$+ \sum_S^{n} \left\{ \frac{y_S}{2B} - \frac{\sin(y_S n\theta/2B)\cos(1-y_s/2B)n\theta}{\sin n\theta} \right\} \frac{m_s}{2B}, \qquad (5a)$$

and

$$P\phi_S = -\sum_1^{S-1} \left\{ \frac{y_s}{2B} - \frac{\sin(y_s n\theta/2B)\cos(1-y_S/2B)n\theta}{\sin n\theta} \right\} f_s$$
$$+ \sum_S^{n-1} \left\{ \left(1-\frac{y_s}{2B}\right) - \frac{\sin(1-y_s/2B)n\theta \cos(y_S n\theta/2B)}{\sin n\theta} \right\} f_s$$
$$+ \sum_0^{S-1} \left\{ 1 - \frac{n\theta \cos(1-y_S/2B)n\theta \cos(y_s n\theta/2B)}{\sin n\theta} \right\} \frac{m_s}{2B}$$
$$+ \sum_S^{n} \left\{ 1 - \frac{n\theta \cos(1-y_s/2B)n\theta \cos(y_S n\theta/2B)}{\sin n\theta} \right\} \frac{m_s}{2B}. \qquad (5b)$$

Restriction 3. For the particular case when $y_s = 2Bs/n$, equations (5*a*) and (5*b*) become

$$\frac{Pw_S}{2B} = \sum_1^{S-1} \left\{ \frac{s}{n}\left(1-\frac{S}{n}\right) - \frac{\sin s\theta \sin(n-S)\theta}{n\theta \sin n\theta} \right\} f_s$$
$$+ \sum_S^{n-1} \left\{ \frac{S}{n}\left(1-\frac{s}{n}\right) - \frac{\sin S\theta \sin(n-s)\theta}{n\theta \sin n\theta} \right\} f_s$$
$$- \sum_0^{S-1} \left\{ \left(1-\frac{S}{n}\right) - \frac{\sin(n-S)\theta \cos s\theta}{\sin n\theta} \right\} \frac{m_s}{2B}$$
$$+ \sum_S^{n} \left\{ \frac{S}{n} - \frac{\sin S\theta \cos(n-s)\theta}{\sin n\theta} \right\} \frac{m_s}{2B}. \qquad (6a)$$

$$P\phi_S = -\sum_1^{S-1} \left\{ \frac{s}{n} - \frac{\sin s\theta \cos(n-S)\theta}{\sin n\theta} \right\} f_s$$
$$+ \sum_S^{n-1} \left\{ \left(1-\frac{s}{n}\right) - \frac{\sin(n-s)\theta \cos S\theta}{\sin n\theta} \right\} f_s$$
$$+ \sum_0^{S-1} \left\{ 1 - \frac{n\theta \cos(n-S)\theta \cos s\theta}{\sin n\theta} \right\} \frac{m_s}{2B}$$
$$+ \sum_S^{n} \left\{ 1 - \frac{n\theta \cos(n-s)\theta \cos S\theta}{\sin n\theta} \right\} \frac{m_s}{2B}. \qquad (6b)$$

3. *Deflection of ribs. Restriction* 4. Assuming that each rib also is of uniform section, the deflections (w) and slopes (ψ) of the ribs in terms of the loads (f) and moments (t) can be obtained from equations (5*a*) and (5*b*) by dividing the right-hand side of each equation by P and finding the limit of each coefficient when $P = 0$. The signs of the terms in f must be changed in order to compensate for change in the relative directions of load and deflection. At the same time y should be changed to x, n to m, S to R, s to r, ϕ to ψ, I_S to I_R, and $2B$ to $2D$.

It is, however, advantageous to defer the derivation of equations relating to ribs from those obtained for stringers until a later stage in the analysis.

Restriction 5. In the special case when $x_R = 2DR/m$ (uniform spacing of stringers), $x_R/2D$ becomes R/m and $x_r/2D$ becomes r/m.

4. *Torsion of ribs and stringers.* Since this analysis is intended to apply only to the *beginning* of buckling, in considering the torsional effects of the moments m_{RS} and t_{RS} we may treat both stringers and ribs as straight, so that the torsional couples act about the axes of the members. Along the S-th rib we have applied torques m_{1S}, m_{2S}, ..., $m_{m-1,S}$, at distances x_1, x_2, ..., x_{m-1} from one end, and at each end torques sufficient to make the twist at each end zero. If μ and μ' are the end torques (positive in the sense opposite to that at m_{1S}, etc.), we have, using restriction 4,

$$CJ_S(\phi_R - \phi_{R-1}) = \Big\{-\mu + \sum_1^{R-1} m_r\Big\}(x_R - x_{R-1}),$$

whence
$$CJ_S\phi_R = -\mu x_R + \sum_1^{R-1} m_r(x_R - x_r).$$

The condition $\phi_m = 0$ gives

$$-2D\mu + \sum_1^{m-1} m_r(2D - x_r) = 0,$$

so that finally

$$CJ_S\phi_R = -\sum_1^{R-1} m_r x_r\Big(1 - \frac{x_R}{2D}\Big) - \sum_R^{m-1} m_r x_R\Big(1 - \frac{x_r}{2D}\Big); \qquad (7a)$$

similarly,

$$CJ_R\psi_S = -\sum_1^{S-1} t_s y_s\Big(1 - \frac{y_S}{2B}\Big) - \sum_S^{n-1} t_s y_S\Big(1 - \frac{y_s}{2B}\Big). \qquad (7b)$$

5. *Particular solution.—(a) General.* The results of §§ 2, 3, and 4 provide all the data necessary for the solution of the problem, but attention

will be confined to the special case when, in addition to the five restrictions already mentioned, the following restrictions are also introduced.

Restriction 6. All the stringers are identical (except the edge stringers).

Restriction 7. All the ribs are identical (except the edge ribs).

These two restrictions result in the following lemma: *any linear expression of the deflections or slopes summed along a rib or stringer is the same function of similar expressions of the forces or moments, similarly summed, as each deflection or slope separately is of the individual forces or moments.*

(*b*) *Summation over stringers.* We now define the following $6n$ functions (σ taking any integral value from 1 to n).

$$ {}_{\sigma}W_R = \sum_{1}^{n-1} w_{Rs} \sin(\sigma\pi s/n), $$

$$ {}_{\sigma}\Phi_R = \sum_{1}^{n-1} \phi_{Rs} \cos(\sigma\pi s/n) + \tfrac{1}{2}(\phi_{R,0} + \phi_{R,n} \cos\sigma\pi)^*, $$

$$ {}_{\sigma}\Psi_R = \sum_{1}^{n-1} \psi_{Rs} \sin(\sigma\pi s/n), \qquad {}_{\sigma}F_R = \sum_{1}^{n-1} f_{Rs} \sin(\sigma\pi s/n), $$

$$ {}_{\sigma}M_R = \sum_{0}^{n} m_{Rs} \cos(\sigma\pi s/n), \qquad {}_{\sigma}T_R = \sum_{1}^{n-1} t_{Rs} \sin(\sigma\pi s/n). $$

We shall henceforth omit the suffix R and the affix σ wherever there is no possibility of confusion.

The coefficient† of f_S in $WP/2B$ is, from equation (6*a*),

$$ \sum_{s}^{n-1} \left\{ \frac{S}{n}\left(1-\frac{s}{n}\right) - \left(\frac{\sin S\theta \sin(n-s)\theta}{n\theta \sin n\theta}\right) \right\} \sin\frac{s\sigma\pi}{n} $$

$$ + \sum_{1}^{S-1} \left\{ \frac{s}{n}\left(1-\frac{S}{n}\right) - \left(\frac{\sin s\theta \sin(n-S)\theta}{n\theta \sin n\theta}\right) \right\} \sin\frac{s\sigma\pi}{n}, $$

* The torsion equation (7*a*) between ϕ_R and m_r is true also for ${}_{\sigma}\Phi_R$ and ${}_{\sigma}M_R$ only if $m_{R,0}$ and $m_{R,n}$ are related to $\frac{1}{2}\phi_{R,0}$ and $\frac{1}{2}\phi_{R,n}$ as m_{Rs} is related to ϕ_{Rs} ($1 \leqslant s \leqslant n-1$). A similar consideration applies in the case of the torsion relation between ψ and t. For this reason a simple solution is possible only if the edge stringers and ribs have half the torsional rigidity of the corresponding main members.

† This coefficient, like the others that follow, can be obtained from the coefficients of equations (6*a*) and (6*b*) by interchanging $\sum_{1}^{S-1}$ with $\sum_{S}^{n-1}$ and S with s. This particular coefficient is not changed by this procedure, because the matrix relating w to f is symmetrical about the principal diagonal.

and it can be shown that this is equal to

$$\frac{(\theta-\sin\theta)\cos\alpha+\sin\theta-\theta\cos\theta}{2n\theta(1-\cos\alpha)(\cos\alpha-\cos\theta)}\sin S\alpha,$$

where $\alpha=\sigma\pi/n$. Thus the terms in the f's in the equation for W reduce to the single term

$$\frac{(\theta-\sin\theta)\cos\alpha+\sin\theta-\theta\cos\theta}{2n\theta(1-\cos\alpha)(\cos\alpha-\cos\theta)}F.$$

Similarly the terms in the m's reduce to the single term

$$\frac{(1-\cos\theta)\sin\alpha}{2(1-\cos\alpha)(\cos\alpha-\cos\theta)}\cdot\frac{M}{2B}.$$

Also we find that

$$P\Phi=\frac{(1-\cos\theta)\sin\alpha}{2(1-\cos\alpha)(\cos\alpha-\cos\theta)}F-\frac{n\theta\sin\theta}{2(\cos\alpha-\cos\theta)}\cdot\frac{M}{2B}$$

and

$$\frac{CJ_S}{2B}\Psi=-\frac{1}{2n(1-\cos\alpha)}T.$$

(*c*) *Summation over ribs.* Clearly a similar process of summation may be applied over the ribs. Moreover, in virtue of the lemma of § (*a*) above, this process will apply not only to the quantities w_{RS}, etc., but also to the derived quantities ${}_\sigma W_R$, etc. (See footnote * on p. 7.) The coefficients of the summations over the ribs may be obtained from those of § (*b*) by putting $\theta=0$, remembering that $P=EI_S(n\theta/2B)^2$, and making the necessary alteration in the sign of the coefficients of F.

We define six new quantities as follows, ρ taking any integral value from 1 to $m-1$:

$${}_{\rho\sigma}W=\sum_1^{m-1}{}_\sigma W_r\sin(\rho\pi r/m), \qquad {}_{\rho\sigma}\Phi=\sum_1^{m-1}{}_\sigma\Phi_r\sin(\rho\pi r/m),$$

$${}_{\rho\sigma}\Psi=\sum_1^{m-1}{}_\sigma\Psi_r\cos(\rho\pi r/m)+\tfrac{1}{2}(\Psi_0+\Psi_m\cos\rho\pi), \qquad {}_{\rho\sigma}F=\sum_1^{m-1}{}_\sigma F_r\sin(\rho\pi r/m),$$

$${}_{\rho\sigma}M=\sum_1^{m-1}{}_\sigma M_r\sin(\rho\pi r/m), \qquad {}_{\rho\sigma}T=\sum_0^{m}{}_\sigma T_r\cos(\rho\pi r/m),$$

and also introduce the following new symbols:

$2b=2B/n=$ spacing of ribs, $2d=2D/m=$ spacing of stringers, $\beta=\pi\rho/m$.

Hence we obtain the following equations:

$$\frac{P}{2b}W=\frac{(\theta-\sin\theta)\cos\alpha+\sin\theta-\theta\cos\theta}{2\theta(1-\cos\alpha)(\cos\alpha-\cos\theta)}F-\frac{(1-\cos\theta)\sin\alpha}{2(1-\cos\alpha)(\cos\alpha-\cos\theta)}\cdot\frac{M}{2b}, \tag{9a}$$

$$P\Phi=\frac{(1-\cos\theta)\sin\alpha}{2(1-\cos\alpha)(\cos\alpha-\cos\theta)}F-\frac{\theta\sin\theta}{2(\cos\alpha-\cos\theta)}\cdot\frac{M}{2b}, \tag{9b}$$

$$\frac{6EI_R}{8d^3}W=\frac{2+\cos\beta}{2(1-\cos\beta)^2}F+\frac{3\sin\beta}{2(1-\cos\beta)^2}\cdot\frac{T}{2d}, \tag{9c}$$

$$\frac{6EI_R}{4d^2}\Psi=\frac{3\sin\beta}{2(1-\cos\beta)^2}F+\frac{3}{1-\cos\beta}\cdot\frac{T}{2d}, \tag{9d}$$

$$\frac{CJ_S}{2b}\Psi=-\frac{1}{2(1-\cos\alpha)}T, \tag{9e}$$

$$\frac{CJ_R}{2d}\Phi=-\frac{1}{2(1-\cos\beta)}M. \tag{9f}$$

These six equations are homogeneous in the six unknowns W, Φ, Ψ, F, M and T. All these variables may therefore be eliminated, leaving one relation between their coefficients. This relation expresses the condition that a solution other than that in which all the variables W, Φ, etc., are zero should be possible; it is therefore a necessary condition for instability. The result of eliminating the six unknowns, W, Φ, etc., from equations (9) is identical with equation (19) at the end of Part II.

If $\sigma=n$, buckling occurs between ribs, the ribs being twisted but not bent and the stringers bent but not twisted. In this case $\alpha=\pi$ and the terms in F and T in equations (9) disappear, so that equations (9b) and (9f) alone lead to the condition for instability

$$\frac{\sin\theta}{\theta(1+\cos\theta)}+\frac{EI_Sd}{CJ_Rb}\cdot\frac{1}{1-\cos\beta}=0. \tag{10}$$

PART II. *Solution by the Rayleigh-Ritz (energy) method.*

1. *Specification of buckled form of grid.* In the unbuckled state all the stringers are considered to be subjected simply to a uniform compressive strain e and the ribs are considered to be unstressed. It is necessary, in addition, to specify only a form for the deflection of the ribs and stringers, out of the original plane of the grid, consequent upon buckling. If we take axes Ox and Oy along one of the edge ribs and one of the edge

stringers respectively, the deflection of the point (x, y) out of the plane is assumed to be represented by the double series

$$w = \sum_r \sum_s \delta_{r,s} \sin \frac{r\pi x}{2D} \sin \frac{s\pi y}{2B},$$

where r and s have all values from 1 to ∞ and where $\delta_{r,s}$ are unknown coefficients.

This form for w is capable of representing any surface over the plane of the grid. Since the deflected centre lines of the ribs and stringers represent only certain sections of this surface, the solution will not lead to a complete specification of the form for w; for instance, it is clear that terms of the type $\delta_{Mm,Nn} \sin (M\pi x/2d) \sin (N\pi y/2b)$, where M and N are any integers, do not affect the deflection of the ribs or stringers. The final form found for w will therefore be, to some extent, indefinite in so far as the coefficients of the type $\delta_{Mm,Nn}$ are concerned.

2. *Flexural energy of stringers and ribs.* If I_S is the moment of inertia of the section of the stringers about the neutral axis parallel to the plane of the grid, the flexural energy of the stringers is

$$\sum_{R=1}^{m-1} \tfrac{1}{2} E I_S \int_0^{2B} \left\{ \frac{\partial^2}{\partial y^2} \left(\sum_r \sum_s \delta_{r,s} \sin \frac{r\pi R}{m} \sin \frac{s\pi y}{2B} \right) \right\}^2 dy$$

$$= \tfrac{1}{2} E I_S \sum_{R=1}^{m-1} \int_0^{2B} \left(\sum_r \sum_s -\frac{s^2\pi^2}{4B^2} \delta_{r,s} \sin \frac{r\pi R}{m} \sin \frac{s\pi y}{2B} \right)^2 dy$$

$$= \tfrac{1}{2} E I_S B \sum_{R=1}^{m-1} \sum_s \frac{s^4\pi^4}{16B^4} \left(\sum_r \delta_{r,s} \sin \frac{r\pi R}{m} \right)^2,$$

since

$$\int_0^{2B} \sin \frac{s_1\pi y}{2B} \sin \frac{s_2\pi y}{2B} \, dy = 0 \text{ if } s_1 \neq s_2, \quad = B \text{ if } s_1 = s_2.$$

Now

$$\sum_{R=1}^{m-1} \sin \frac{R\pi r_1}{m} \sin \frac{R\pi r_2}{m} = \pm \tfrac{1}{2} m \text{ if } r_2 = 2Mm \pm r_1, \text{ where } M \text{ is any integer (including zero) and values of } r_1 \text{ and } r_2 \text{ multiples of } m \text{ are excluded,}$$

$$= 0 \text{ if } r_2 \neq 2Mm \pm r_1, \text{ or if } r_1 \text{ and } r_2 \text{ are multiples of } m.$$

The total flexural strain energy of the stringers is therefore

$$(\pi^4 mEI_S/64B^3)\sum_s s^4 \sum_r{}'(\delta_{r,s}-\delta_{2m-r,s}+\delta_{2m+r,s}-\dots)^2,$$

$$\left(\text{where } \sum_r{}' \text{ denotes summation for all values of } r \text{ less than } m\right)^*$$

$$= G\sum_s\sum_r{}' s^4({}_r\Delta_s)^2,$$

where $G = \pi^4 mEI_S/64B^3$, ${}_r\Delta_s = \delta_{r,s}-\delta_{2m-r,s}+\delta_{2m+r,s}-\dots$.

Similarly, the flexural energy of the ribs is $H\sum_r\sum_s{}' r^4({}_s\Delta_r)^2$, where $H = \pi^4 nEI_R/64D^3$, I_R is the moment of inertia of the rib section, $\sum_s{}'$ denotes summation for all values of s less than n, and

$${}_s\Delta_r = \delta_{r,s}-\delta_{r,2n-s}+\delta_{r,2n+s}-\dots.$$

4. *Direct stress energy of the stringers.* The shortening of the R-th stringer per unit length due to buckling is

$$\frac{1}{4B}\int_0^{2B}\left(\frac{\partial}{\partial y}\sum_r\sum_s \delta_{r,s}\sin\frac{r\pi R}{m}\sin\frac{s\pi y}{2B}\right)^2 dy = \frac{\pi^2}{16B^2}\sum_s s^2\left(\sum_r \delta_{r,s}\sin\frac{R\pi r}{m}\right)^2.$$

The total direct stress energy stored in the stringers (omitting the edge stringers) is therefore

$$\tfrac{1}{2}EA_S\,2B\sum_{R=1}^{m-1}\left\{e-\frac{\pi^2}{16B^2}\sum_s s^2\left(\sum_r \delta_{r,s}\sin\frac{R\pi r}{m}\right)^2\right\}^2$$

(where A_S is the area of the stringer section)

$$= EA_S B\left[(m-1)e^2 - e\,\frac{m\pi^2}{16B^2}\sum_s\sum_r{}' s^2({}_r\Delta_s)^2 + (\text{term in } \delta^4_{r,s})\right]$$

$$= G\left[(\text{term in } e^2) - \frac{e}{e_s}\,n^2\sum_s\sum_r{}' s^2({}_r\Delta_s)^2 + (\text{term in } \delta^4_{r,s})\right],$$

where $e_s = \pi^2 I_S/4b^2A_S$ is the buckling strain of a stringer simply supported between adjacent ribs.

5. *Torsional energy of stringers and ribs.* The torsional strain energy of the R-th stringer is

$$\tfrac{1}{2}CJ_S{}'\int_0^{2B}\left\{\frac{\partial^2}{\partial x\,\partial y}\left(\sum_r\sum_s \delta_{r,s}\sin\frac{r\pi x}{2D}\sin\frac{s\pi y}{2B}\right)\right\}^2 dy,$$

* All possible values of r are thus included. For, if r is restricted to values less than m, the series δ_r, δ_{2m-r}, ..., etc. include all values of the suffix that are not multiples of m.

where J_S' is the polar moment of inertia of the stringer section, which is J_S for $R=1$ to $m-1$ inclusive, and $\frac{1}{2}J_S$ for $R=0$ and $R=m$; C is the torsion modulus; and $x=2DR/m$. The total torsional strain energy of the stringers is therefore

$$\tfrac{1}{2}CJ_S\left[\tfrac{1}{2}\int_0^{2B}\left(\sum_r\sum_s \frac{\pi^2 rs}{4BD}\,\delta_{r,s}\cos\frac{s\pi y}{2B}\right)^2 dy + \sum_{R=1}^{m-1}\int_0^{2B}\left(\sum_r\sum_s \frac{\pi^2 rs}{4BD}\,\delta_{r,s}\cos\frac{R\pi r}{m}\cos\frac{s\pi y}{2B}\right)^2 dy + \tfrac{1}{2}\int_0^{2B}\left(\sum_r\sum_s \frac{\pi^2 rs}{4BD}\,\delta_{r,s}\cos r\pi\cos\frac{s\pi y}{2B}\right)^2 dy\right]$$

$$=\frac{J}{m}\sum_s s^2\left\{\left(\sum_r r\delta_{r,s}\right)^2+2\sum_{R=1}^{m-1}\left(\sum_r r\delta_{r,s}\cos\frac{R\pi r}{m}\right)^2+\left(\sum_r r\delta_{r,s}\cos r\pi\right)^2\right\},$$

where $J=\pi^4 mCJ_S/64BD^2$. Now

$$\sum_{R=1}^{m-1}\cos\frac{R\pi r_1}{m}\cos\frac{R\pi r_2}{m}=\tfrac{1}{2}(m-2)$$ if $r_2=2Mm\pm r_1$, where M is any integer (including zero) and values of r_1 and r_2 multiples of m are excluded,

$=m-1$ if r_1 and r_2 are both multiples of m,

$=0$ if $r_2\neq 2Mm\pm r_1$ and r_1 and r_2 are one odd and one even,

$=-1$ if $r_2\neq 2Mm\pm r_1$ and r_1 and r_2 are both odd or both even.

As a result of these relations the torsional strain energy of the stringers becomes

$$J\Sigma s^2\left\{\tfrac{1}{2}({}_0\Delta_s')^2+\sum_r{}'({}_r\Delta_s')^2+\tfrac{1}{2}({}_m\Delta_s')^2\right\},$$

where

$${}_r\Delta_s'=r\delta_{r,s}+(2m-r)\,\delta_{2m-r,s}+(2m+r)\,\delta_{2m+r,s}+\ldots.$$

It should be remarked that

$${}_m\Delta_s'=m\delta_{m,s}+(2m-m)\,\delta_{2m-m,s}+\ldots=2(m\delta_{m,s}+3m\delta_{3m,s}+\ldots)$$

and

$${}_0\Delta_s'=2(2m\delta_{2m,s}+4m\delta_{4m,s}+\ldots).$$

The factor $\frac{1}{2}$ multiplying the terms in ${}_m\Delta_s'$ and ${}_0\Delta_s'$ is really $2\times\frac{1}{4}$, the factor 2 being introduced by the R summation and the factor $\frac{1}{4}$ because each term of the m and $2m$ series is, in effect, counted twice.

The torsional strain energy of the ribs can be expressed in a similar form.

6. *Total strain energy.* Omitting terms in δ^4, which may be neglected in comparison with those in δ^2 when considering the *beginning* of buckling, and terms not containing the δ's, the total strain energy is

$$G\sum_s\sum_r{}' s^2(s^2-n^2e/e_s)({}_r\Delta_s)^2+H\sum_r\sum_s{}' r^4({}_s\Delta_r)^2$$
$$+J\sum_s s^2\{\tfrac{1}{2}({}_0\Delta_s')^2+\sum_r{}'({}_r\Delta_s')^2+\tfrac{1}{2}({}_m\Delta_s')^2\}$$
$$+K\sum_r r^2\{\tfrac{1}{2}({}_0\Delta_r')^2+\sum_s{}'({}_s\Delta_r')^2+\tfrac{1}{2}({}_n\Delta_r')^2\},$$

where $G=\pi^4mEI_S/64B^3$, $H=\pi^4nEI_R/64D^3$,

$J=\pi^4mCJ_S/64BD^2$, $K=\pi^4nCJ_R/64B^2D$,

and

$${}_r\Delta_s=\delta_{r,s}-\delta_{2m-r,s}+\delta_{2m+r,s}-\cdots,$$
$${}_s\Delta_r=\delta_{r,s}-\delta_{r,2n-s}+\delta_{r,2n+s}-\cdots,$$
$${}_r\Delta_s'=r\delta_{r,s}+(2m-r)\,\delta_{2m-r,s}+(2m+r)\,\delta_{2m+r,s}+\cdots,$$
$${}_s\Delta_r'=s\delta_{r,s}+(2n-s)\,\delta_{r,2n-s}+(2n+s)\,\delta_{r,2n+s}+\cdots,$$
$$\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots$$

It will be seen that, in the above expression for the total strain energy, the coefficients δ occur in groups of the type

$\delta_{r,s}$	$\delta_{2m-r,s}$	$\delta_{2m+r,s}$	$\delta_{4m-r,s}$	$\cdots$
$\delta_{r,2n-s}$	$\delta_{2m-r,2n-s}$	$\delta_{2m+r,2n-s}$	$\cdots$	$\cdots$
$\delta_{r,2n+s}$	$\delta_{2m-r,2n+s}$	$\delta_{2m+r,2n+s}$	$\cdots$	$\cdots$
$\cdots$ $\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

the general term in any group with the dominant term $\delta_{r,s}$ ($r<m$ and $s<n$)* being $\delta_{2Mm\pm r,2Nn\pm s}$, where M and N are any integers. Each δ occurs in one group and one group only; in that group it occurs as a cross product with itself and with each other member of the group, and no cross products

* The groups for which $r=m$ or $2m(0)$ or $s=n$ or $2n(0)$ will be discussed separately later.

of members of different groups occur. Thus, if the conditions for minimum energy are formed by equating to zero the differentials of the expression for the energy with respect to each δ, the resulting equations (linear and homogeneous in the δ's) are grouped in sets such that any one set involves only the δ's belonging to one group. Since the number of equations in a set is equal to the number of δ's in the corresponding group, either all the δ's are zero or an equation of compatibility between the coefficients must be satisfied. It is this equation of compatibility, expressing the condition that a solution other than that in which all the δ's are zero should be possible, which may be termed the equation of buckling. Naturally this equation must involve the strain e.

Now it is clear that, to each group represented by its dominant term δ_{rs} $(r < m,\ s < n)$, there corresponds a different equation of compatibility, and in general the values of e satisfying any one of these equations will not simultaneously satisfy any of the others*. Thus, if a non-zero solution in the δ's of one group is possible, all the δ's of all the other groups must be identically zero. Now $\sin(2Mm \pm r)\theta$ has, at the values $\theta = \pi R/m$, values proportional to $\sin r\theta$ simply. Therefore at the points of intersection of stringers and ribs, all the members of a group of δ's give deflections proportional to those of the dominant term. In other words, in each possible buckled form all the points of intersection of the stringers and ribs lie on a simple sine surface of the type

$$\delta_{r,s} \sin(r\pi x/2D) \sin(s\pi y/2B),$$

and the subordinate terms in the δ group serve only to modify the form of the stringers and ribs between points of intersection.

Groups for which $r = m$ or $2m$ make the coefficient of G in the expression for the total energy zero. Therefore the equations of the compatibility for these groups do not contain e and no instability is possible. Groups for which $r = m$ or $2m$ and $s = n$ or $2n$ involve only the terms in J and K. The internal energy due to these groups is therefore identically zero if ${}_n\Delta'_{Mm} = 0$ for all values of M and ${}_m\Delta'_{Nn} = 0$ for all values of N. It is obvious that these equations merely express the conditions that the twist of the ribs and stringers should be everywhere zero, but it should be noticed that, since Δ's are unrestricted, solutions not all zero in the δ's are possible. Groups for which $s = n$ or $2n$ and $r < m$ make the coefficient of H zero; these cases are not discussed in the

* Special geometry of the grid may result in the simultaneous satisfaction of two or more equations of compatibility by the same value or values of e, but such special cases are here excluded.

present paper, but may be shown to lead to the same result as that obtained in Part I by the exact method, equation (10).

7. *Solution of the equations for minimum strain energy.* Differentiation of the total energy with respect to $\delta_{r,s}$, $\delta_{2m-r,s}$, ..., $\delta_{r,2n-s}$, $\delta_{2m-r,2n-s}$, ..., in succession yields the equations

$$\pm Gs^2(s^2-n^2e/e_s)\Delta_s+H(2Mm\pm r)^4\Delta_{2Mm\pm r}$$
$$+Js^2(2Mm\pm r)\Delta_s'+K(2Mm\pm r)^2 s\,\Delta'_{2Mm\pm r}=0 \quad (11a)$$

$$\pm G(2Nn-s)^2\{(2Nn-s)^2-n^2e/e_s\}\Delta_{2Nn-s}-H(2Mm\pm r)^4\Delta_{2Mm\pm r}$$
$$+J(2Nn-s)^2(2Mm\pm r)\Delta'_{2Nn-s}+K(2Mm\pm r)^2(2Nn-s)\Delta'_{2Mm\pm r}=0, \quad (11b)$$

and

$$\pm G(2Nn+s)^2\{(2Nn+s)^2-n^2e/e_s\}\Delta_{2Nn+s}+H(2Mm\pm r)^4\Delta_{2Mm\pm r}$$
$$+J(2Nn+s)^2(2Mm\pm r)\Delta'_{2Nn+s}+K(2Mm\pm r)^2(2Nn+s)\Delta'_{2Mm\pm r}=0, \quad (11c)$$

where Δ_s is written for ${}_r\Delta_s$, Δ_r for ${}_s\Delta_r$, ..., and M and N can have any integral values (including zero only for positive r or s). The equations are written in three groups to avoid confusion about signs.

In each group of constant s, by adding or subtracting the first equation ($M=0$) from each of the succeeding equations in the same group, we can eliminate the terms in G, *i.e.*, in Δ_s, Δ_{2n-s}, ... The equations then become

$$H\{(2Mm\pm r)^4\Delta_{2Mn\pm r}\mp r^4\Delta_r\}+J(2Nn+s)^2\cdot 2Mm\Delta'_{2Nn+s}$$
$$+K(2Nn+s)\{(2Mm\pm r)^2\Delta'_{2Mm\pm r}\mp r^2\Delta_r'\}=0 \quad (12a)$$

and

$$-H\{(2Mm\pm r)^4\Delta_{2Mm\pm r}\mp r^4\Delta_r\}+J(2Nn-s)^2\,2Mm\Delta'_{2Nn-s}$$
$$+K(2Nn-s)\{(2Mm\pm r)^2\Delta'_{2Mm\pm r}\mp r^2\Delta_r'\}=0. \quad (12b)$$

By adding the equation in s ($N=0$) to the equation in $2Nn-s$ (for each value of r) and subtracting it from the equation in $2Nn+s$, we can eliminate the term in H, leaving the system of equations

$$2Mm\cdot J\{(2Nn\pm s)^2\Delta'_{2Nn\pm s}\mp s^2\Delta_s'\}$$
$$+2NnK\{(2Mm\pm r)^2\Delta'_{2Mm\pm r}\mp r^2\Delta_r'\}=0 \quad (13)$$

for all integral values of M and N (zero excluded). These equations can be satisfied only if

$$(2Mm \pm r)^2 \Delta'_{2Mm \pm r} \mp r^2 \Delta_r' = 2Mm \,.\, P,$$

and

$$(2Nn \pm s)^2 \Delta'_{2Nn \pm s} \mp s^2 \Delta_s' = 2Nn \,.\, Q,$$

where P and Q are independent of M and N, and $JQ + KP = 0$. Hence we find also

$$H\{(2Mm \pm r)^4 \Delta_{2Mm \pm r} \mp r^4 \Delta_r\} = -2Mm(Js^2 \Delta_s' + KsP) \qquad (14)$$

and

$$G(2Nn \pm s)^2 \{(2Nn \pm s)^2 - n^2 e/e_s\} \Delta_{2Nn \pm s}$$
$$= \mp Hr^4 \Delta_r - Jr(2NnQ \pm s^2 \Delta_s') - Kr^2(2Nn \pm s) \Delta_r'. \qquad (15)$$

By these relations the number of unknowns is reduced to four, namely Δ_r, Δ_r', Δ_s' and P (or Q). We have also four identities between the Δ's, namely

$$\Delta_r - \Delta_{2m-r} + \Delta_{2m+r} - \ldots = \Delta_s - \Delta_{2n-s} + \Delta_{2n+s} - \ldots,$$

$$r\Delta_r + (2m-r)\Delta_{2m-r} + (2m+r)\Delta_{2m+r} + \ldots = \Delta_s' - \Delta'_{2n-s} + \Delta'_{2n+s} - \ldots$$

$$\Delta_r' - \Delta'_{2m-r} + \Delta'_{2m+r} - \ldots = s\Delta_s + (2n-s)\Delta_{2n-s} + (2n+s)\Delta_{2n+s} + \ldots$$

and

$$r\Delta_r' + (2m-r)\Delta'_{2m-r} + \ldots = s\Delta_s' + (2n-s)\Delta'_{2n-s} + \ldots.$$

By substitution in these identities for Δ_{2m-r}, ..., in terms of Δ_r, Δ_r', Δ_s' and P, four linear equations in the four unknowns are obtained. Since these equations are homogeneous, the unknowns may be eliminated, leaving one equation, which is, in fact, the equation of buckling. If this equation is not satisfied the only possible solution is all the δ's (and Δ's) zero.

By summation of the series involved in the expressions

$$r\Delta_r' + (2m-r)\Delta'_{2m-r} + \ldots \quad \text{and} \quad s\Delta_s' + (2n-s)\Delta'_{2n-s} + \ldots,$$

it may be shown that the coefficients of P and Q in these sums are infinite. Since at the commencement of buckling all the δ's and hence all the Δ's may be regarded as arbitrarily small, the sums $r\Delta_r' + (2m-r)\Delta'_{2m-r} + \ldots$ must be finite. Thus P must be zero.

The remaining series involved in the necessary summations are of the forms

$$\sum_N \left[\frac{1}{(s\pm 2Nn)^2-a^2}\right], \quad \sum_N \left[\frac{1}{(s\pm 2Nn)\{(s\pm 2Nn)^2-a^2\}}\right],$$

$$\sum_N \left[\frac{1}{(s\pm 2Nn)^2\{(s\pm 2Nn)^2-a^2\}}\right].$$

The sums of these series are

$$\frac{\pi}{2na}\cdot\frac{\sin \pi a/n}{\cos(\pi a/n)-\cos(\pi s/n)}, \quad \frac{\pi}{2na^2}\cdot\frac{\sin(\pi s/n)\{1-\cos(\pi a/n)\}}{\{\cos(\pi a/n)-\cos(\pi s/n)\}\{1-\cos(\pi s/n)\}}.$$

and

$$\frac{\pi}{2na^3}\cdot\frac{\sin(\pi a/n)-(\pi a/n)\cos(\pi a/n)+\{(\pi a/n)-\sin(\pi a/n)\}\cos(\pi s/n)}{\{\cos(\pi a/n)-\cos(\pi s/n)\}\{1-\cos(\pi s/n)\}},$$

respectively.

The first three identities then take the forms

$$\frac{\sin\theta-\theta\cos\theta+(\theta-\sin\theta)\cos\alpha}{\theta^3(\cos\alpha-\cos\theta)(1-\cos\alpha)}\cdot\frac{H}{G}\cdot X+\frac{\sin\alpha(1-\cos\theta)}{\theta^2\alpha(\cos\alpha-\cos\theta)(1-\cos\alpha)}\cdot\frac{K}{G}\cdot Z$$

$$=\left(\frac{n}{m}\right)^4\left[\frac{2+\cos\beta}{6(1-\cos\beta)^2}X-\frac{\sin\beta}{2\beta(1-\cos\beta)^2}Y\right]. \quad (16)$$

$$\frac{H}{J}\cdot\frac{1}{1-\cos\alpha}\cdot Y=\left(\frac{n}{m}\right)^2\left[\frac{\beta\sin\beta}{2(1-\cos\beta)^2}X-\frac{1}{1-\cos\beta}Y\right], \quad (17)$$

$$\frac{\alpha\sin\alpha(1-\cos\theta)}{\theta^2(\cos\alpha-\cos\theta)(1-\cos\alpha)}\cdot\frac{H}{G}\cdot X+\frac{\sin\theta}{\theta(\cos\alpha-\cos\theta)}\cdot\frac{K}{G}\cdot Z$$

$$=\left(\frac{n}{m}\right)^2\cdot\frac{1}{1-\cos\beta}\cdot Z, \quad (18)$$

where

$$X=r^4\Delta_r+\frac{J}{H}rs^2\Delta_s', \qquad \theta=\pi\sqrt{(e/e_s)},$$

$$Y=\frac{J}{H}rs^2\Delta_s', \qquad \alpha=\pi s/n,$$

$$Z=r^2s\Delta_r', \qquad \beta=\pi r/m.$$

By eliminating X, Y and Z from these equations, the condition for instability is found in the form

$$\frac{\sin\theta-\theta\cos\theta+(\theta-\sin\theta)\cos\alpha}{\theta^3(\cos\alpha-\cos\theta)(1-\cos\alpha)} - \frac{(1+\cos\alpha)(1-\cos\theta)^2}{\theta^4(\cos\alpha-\cos\theta)^2(1-\cos\alpha)}\Big/\left\{\frac{\sin\theta}{\theta(\cos\alpha-\cos\theta)}-\frac{n^2}{m^2}\cdot\frac{1}{1-\cos\beta}\cdot\frac{G}{K}\right\}$$

$$=\frac{n^4}{m^4}\frac{G}{H}\left[\frac{2+\cos\beta}{6(1-\cos\beta)^2}-\frac{1+\cos\beta}{4(1-\cos\beta)^3}\Big/\left\{\frac{1}{1-\cos\beta}+\frac{m^2}{n^2}\cdot\frac{1}{1-\cos\alpha}\cdot\frac{H}{J}\right\}\right], \quad (19)$$

which is identical with that found by the exact method in Part I.

Part III. *Discussion of results.*

1. *General interpretation of results.* The solutions in terms of θ of equation (19) for different pairs of values of α and β define a series of critical values of $e=(\theta/\pi)^2 e_s$ at which the grid would buckle into each of the forms defined by the values of σ and ρ corresponding to the given values of α and β. In the buckled form, the points of intersection of the stringers and ribs lie on the surface $w=A\sin(\rho\pi x/2D)\sin(\sigma\pi y/2B)$, which has ρ half waves across the width and σ half waves down the length of the grid. The equation in θ for given values of α and β may have more than one root*, but for practical purposes only the lowest root need be considered.

For practical purposes also attention could be restricted to buckling in the *form* corresponding to the lowest value of θ, but, unfortunately, it is not generally possible to decide which form this is, for the form of equation (19) is so complex that the conditions for minimum θ for variations of α and β cannot be expressed directly. It is, however, possible to examine the practical problem indirectly by regarding equation (19) in a slightly different way.

2. *Practical interpretation of results.* If the values of α, β, m^2K/n^2G and n^2J/m^2H are given, equation (19) can be used to determine the values of m^4H/n^4G for any given value of θ. For this purpose it is

* *E.g.*, if K is negligible, the $\alpha=\pi$ ($\sigma=n$) solution gives the values $\theta=\pi, 3\pi, 5\pi, \ldots$, corresponding to the odd solutions of the ordinary simply supported strut problem. The even solutions $2\pi, 4\pi, \ldots$ are given by the case $\alpha=2\pi$ $\sigma=2n$ or zero).

convenient to write

$$p = 1-\cos\alpha, \qquad q = 1-\cos\beta,$$

$$\xi = \frac{m^2 K}{n^2 G}\left(=\frac{bCJ_R}{dEI_S}\right), \qquad \eta = \frac{n^2 J}{m^2 H}\left(=\frac{dCJ_S}{bEI_R}\right).$$

Equation (19) then becomes

$$\frac{m^4}{n^4}\cdot\frac{H}{G}$$

$$= \frac{6-2q+\eta p}{12q(q+\eta p)}\cdot\frac{\theta^2 p\{(1-\cos\theta)-p-\xi q\sin\theta/\theta\}}{(1-\cos\theta)-p(1-\sin\theta/\theta)+\xi q\{2(1-\cos\theta)/\theta^2-\sin\theta/\theta\}} \qquad (20)$$

and equation (10) becomes

$$\theta(1+\cos\theta)/\sin\theta = -\xi q. \qquad (21)$$

Whatever be the value of H/G, the ultimate strength of the grid cannot exceed that corresponding to the lowest root θ_1 of equation (21), for that root determines the occurrence of buckling between ribs, and this form of buckling is not affected by the flexural rigidity of the ribs. Therefore the practical objective is so to choose the ratio H/G that the lowest root θ_2 of equation (20) for all possible values of p and q should be equal to θ_1. For, if $\theta_1 > \theta_2$, some of the potential strength of the stringers is sacrificed, whereas, if $\theta_2 > \theta_1$, the ribs are stiffer than they need be.

It can be shown that the highest value of H/G determined from equation (20) by variation of q only is always given by the lowest value of q, viz. $1-\cos\pi/m$, corresponding to $\rho = 1$. This represents a single half-wave across the width of the grid. On the other hand, the value of p which makes H/G a maximum depends upon the value of θ. In general, each particular case for given values of ξ, η, and θ must be considered separately; but, if $\theta = \theta_1$, equation (20) becomes

$$\frac{m^4 H}{n^4 G} = \frac{6-2q+\eta p}{12q(q+\eta p)}\cdot\frac{\theta_1^2 p}{1-(\sin\theta_1)/\theta_1}. \qquad (22)$$

Since $6-2q$, q, and η are all positive, H/G increases continuously with p, so that the maximum value of H/G is given by the greatest possible value of p. This value is $p = 1+\cos(\pi/n)$, given by $\sigma = n-1$; the form of buckling is such that the number of half-waves down the length of the grid is one less than the number of sections into which the stringers are divided by the ribs.

In practice, considerable interest is attached to the case when ξ and η are small and m and n are large. In this case $p = 2$, $q = \pi^2/2m^2$, and $\theta_1 = \pi + 2\xi q/\pi$ approximately. Insertion of these values in equation (22) gives

$$\frac{m^4 H}{n^4 G} = \frac{4m^4}{\pi^2}\left(1 + \frac{\xi}{m} - \frac{4}{\pi^2} m^2 \eta\right),$$

which can be expressed in the more convenient form

$$\frac{H}{G} = \frac{4}{\pi^2} n^2 \left(n^2 + \frac{K - J}{G}\right). \tag{23}$$

Either equation (22) or (23) is suitable for design purposes; it appears from the latter equation that if the torsional rigidities are small they have practically no effect on the ratio H/G.

3. *Buckling of a grid without torsion of its members.* If the stringers and ribs have no torsional rigidity, or if the stringers are connected to the ribs by universal joints, equation (19) reduces to the form

$$\frac{m^4 H}{n^4 G} = \frac{2 + \cos\beta}{6(1 - \cos\beta)^2} \cdot \frac{\theta^3(\cos\alpha - \cos\theta)(1 - \cos\alpha)}{\sin\theta - \theta\cos\theta + (\theta - \sin\theta)\cos\alpha} \tag{24a}$$

or

$$\frac{I_R}{I_S} \cdot \frac{b^3}{d^3} \cdot \frac{6(1 - \cos\beta)^2}{2 + \cos\beta} = \frac{\theta^3(\cos\alpha - \cos\theta)(1 - \cos\alpha)}{\sin\theta - \theta\cos\theta + (\theta - \sin\theta)\cos\alpha}. \tag{24b}$$

The value of α defines the form of buckling, and the value of the stiffness ratio $(I_R b^3/I_S d^3)$ can be plotted against the load (proportional to θ^2) for each possible value of α. The type of diagram thus obtained is illustrated by Fig. 2, which is drawn for $n = 6$. At any load less than $P_0 s^2/n^2$, where P_0 is the inter-rib buckling load, there is no liability to buckle in any mode for which $\sigma \geqslant s$; but some rib stiffening is required to prevent buckling in each of the modes for which $\sigma < s$. In general, if $s_2 > s_1$, the stiffness ratio necessary to prevent s_2 buckling becomes greater than that required to prevent s_1 buckling when

$$\frac{\theta}{\sin\theta} \cdot \frac{\{\cos\theta - \cos(\pi s_1/n)\}\{\cos\theta - \cos(\pi s_2/n)\}}{\{1 - \cos(\pi s_1/n)\}\{1 - \cos(\pi s_2/n)\}} > 1.$$

Each separate line in Fig. 2 is defined by one value of σ. The diagram for $n = 6$ is therefore applicable to the case $n = 3$ if the lines for $\sigma/n = \frac{1}{6}$, $\frac{3}{6}$ and $\frac{5}{6}$ are omitted, or to the case $n = 12$ if lines for $\sigma/n = \frac{1}{12}$, $\frac{3}{12}$, $\frac{5}{12}$, $\frac{7}{12}$, $\frac{9}{12}$, and $\frac{11}{12}$ are added. If n is large, σ/n can assume any value. In this case it can be shown that the greatest value of the right-hand

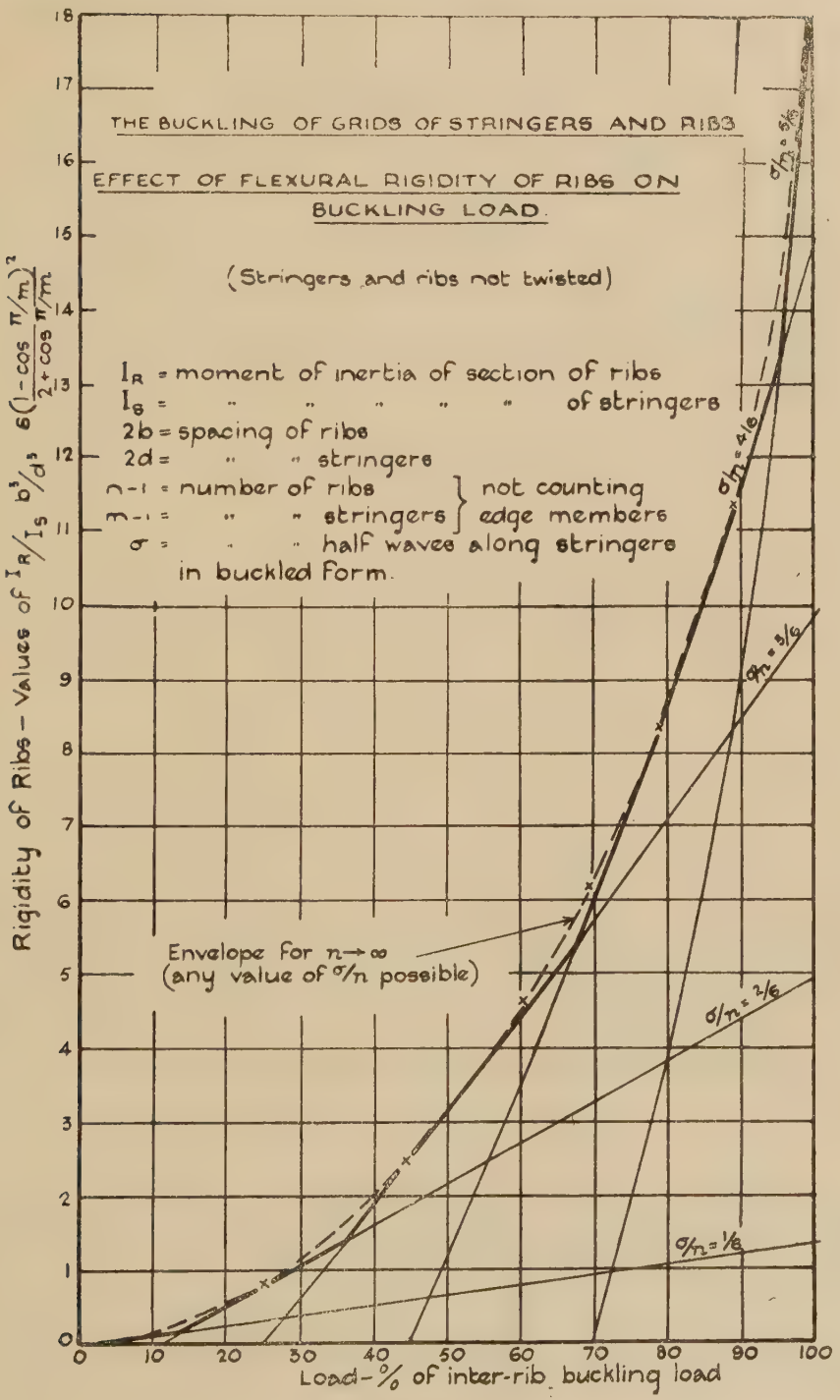

THE BUCKLING OF GRIDS OF STRINGERS AND RIBS
EFFECT OF FLEXURAL RIGIDITY OF RIBS ON BUCKLING LOAD.
(Stringers and ribs not twisted)
I_R = moment of inertia of section of ribs
I_S = " " " " " of stringers
2b = spacing of ribs
2d = " " stringers
n-1 = number of ribs } not counting
m-1 = " " stringers } edge members
σ = " " half waves along stringers in buckled form.
Rigidity of Ribs – Values of $\frac{I_R}{I_S}\frac{b^3}{d^3}\frac{6\left(\frac{1-\cos\pi/m}{2+\cos\pi/m}\right)^2}{\pi/m}$
Load – % of inter-rib buckling load
Envelope for n→∞ (any value of σ/n possible)
σ/n = 6/6
σ/n = 4/6
σ/n = 3/6
σ/n = 2/6
σ/n = 1/6
0 10 20 30 40 50 60 70 80 90 100
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Fig. 2.

side of equation (24) occurs when

$$\cos\alpha = \frac{\cos\theta + \sqrt{(\sin\theta/\theta)}}{1+\sqrt{(\sin\theta/\theta)}},$$

and the value of this maximum is

$$\frac{\theta^2(1-\cos\theta)}{\{1+\sqrt{(\sin\theta/\theta)}\}^2}.$$

It should be noted that as θ approaches π, $\cos\alpha$ approaches the value of -1, and σ approaches n.

This limit for $n\to\infty$ gives the inner envelope shown dotted in Fig. 2. It is also worth noting that

$$\frac{\theta^2(1-\cos\theta)}{\{1+\sqrt{(\sin\theta/\theta)}\}^2} = \frac{\theta^4}{8}\left(1+\frac{\theta^4}{2880}+\frac{\theta^6}{120960}+\ldots\right),$$

so that for values of θ up to about 0·8 (*i.e.* 64 per cent. of the inter-rib buckling load) the required stiffness ratio is very nearly equal to $\theta^4/8$; at $\theta=\pi$, however, the required ratio is $2\pi^2$, which is greater than $\pi^4/8$ in the ratio $16/\pi^2$. This point is better illustrated in Fig. 3, where

$$\left\{\frac{I_R}{I_S}\frac{b^3}{d^3}\frac{6(1-\cos\beta)^2}{2+\cos\beta}\right\}^{\frac{1}{2}}$$

is shown plotted against $100(\theta/\pi)^2$.

4. *Comparison with approximate solutions.* If, in Part II, r and s are restricted to values less than m and n respectively, in each group of δ's every member except the dominant term disappears. The wave form thus defined by a single δ_{rs} still correctly describes a possible configuration of the lattice points of the grid, but the deflections of the stringers and ribs between the lattice points are not properly represented*. This approximate representation of the buckled form leads to the buckling equation

$$n^2\left(\frac{\theta}{\pi}\right)^2 = n^2\frac{e}{e_s} = \sigma^2+\frac{H}{G}\frac{1}{\sigma^2}+\frac{K+J}{G}. \qquad (25)$$

The minimum value of θ is given by $\sigma^2=\sqrt{(H/G)}$, where σ is the number of waves down the length of the grid, so that, if σ can assume any value,

* For there is no discontinuity in any of the differentials of the deflection at the lattice points, so that this solution ignores the possibility of interchange of load between stringers and ribs. The authors are indebted to Prof. Southwell for drawing their attention to this error in the approximate solution.

equation (25) reduces to

$$n^2\left(\frac{\theta}{\pi}\right)^2 = 2\sqrt{\left(\frac{H}{G}\right)} + \frac{K+J}{G}. \tag{26}$$

This equation is identical with that previously obtained by J. L. Taylor* by treating the stiffnesses of the stringers and ribs as distributed over the whole plane of the grid. This correspondence between the two approximate solutions is due to the fact that the deflected form of a uniform beam under a distributed load proportional to the deflection is a simple sine curve; this correspondence also illustrates the reason why the edge members of the grid must have half the torsional rigidity of the main members.

Direct comparison between equations (20) and (25) is not practicable; if, however, attention is restricted to the value

$$\theta = \theta_1 = \pi + \frac{2\xi q}{\pi} = \pi\left(1 + \frac{K}{n^2 G}\right),$$

which represents approximately the load at which buckling between the ribs occurs, equation (26) becomes

$$n^2\left(1 + \frac{K}{n^2 G}\right)^2 = 2\sqrt{\left(\frac{H}{G}\right)} + \frac{K+J}{G},$$

which, to the same degree of approximation, may be written

$$\frac{H}{G} = \frac{n^2}{4}\left(n^2 + \frac{K-J}{G}\right). \tag{27}$$

Equation (27) is similar to equation (23), obtained by approximation from the exact solution, except that the constant $4/\pi^2$ is reduced in the ratio $\pi^2/16$. This is the same ratio as that encountered at the end of § 3 above.

If the torsional rigidities are neglected, equation (26) becomes

$$\sqrt{\left(\frac{H}{G}\right)} = \frac{n^2}{2}\left(\frac{\theta}{\pi}\right)^2,$$

to which equation (24) also approximates if m is large and θ is not very near π. The approximate solution is shown plotted on a diagram similar to Fig. 3 in Fig. 4. The envelope for indefinitely large n is, in this

* J. L. Taylor, "Stability of a monocoque in compression", *Aero. Res. Committee, R. and M.*, no. 1679.

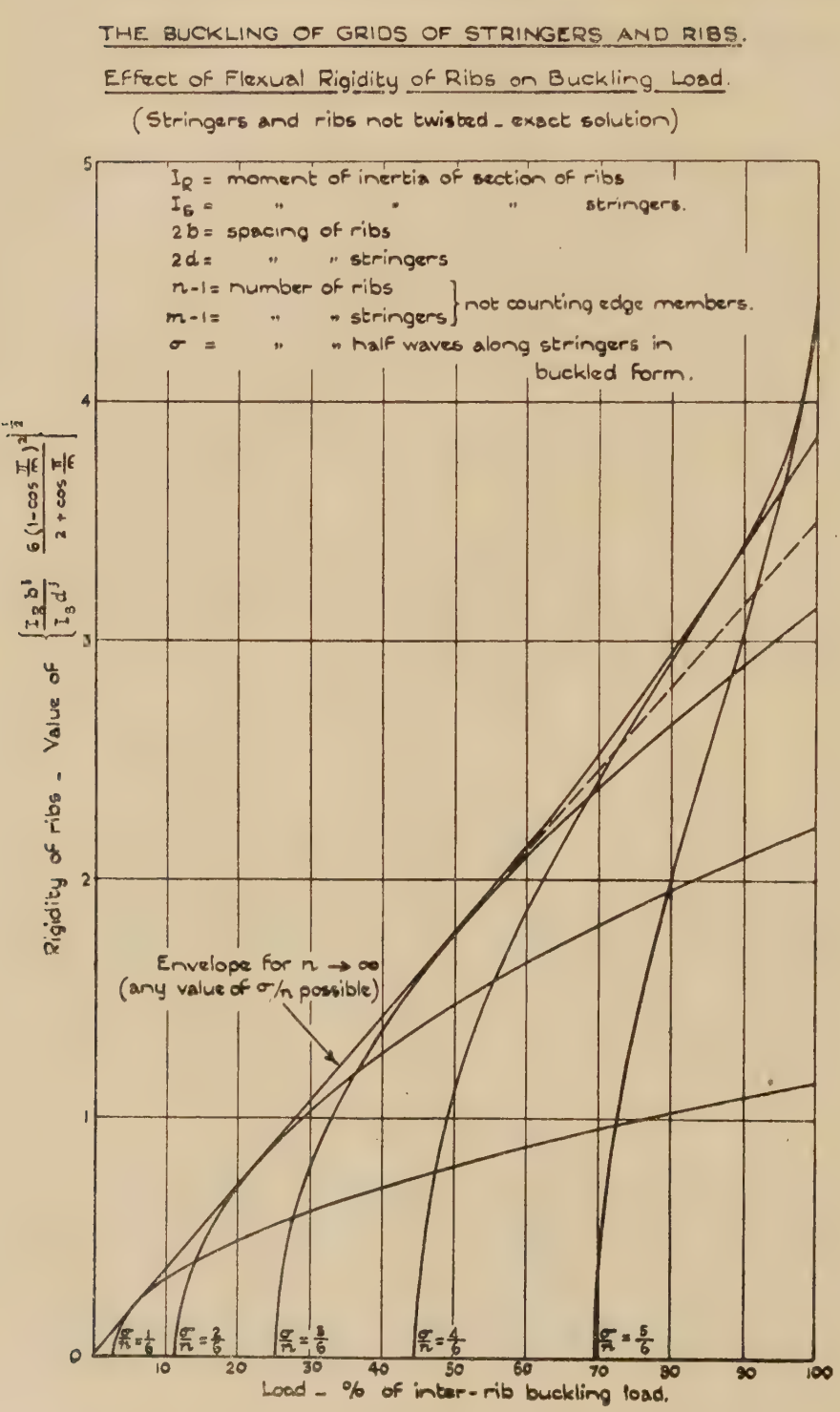

THE BUCKLING OF GRIDS OF STRINGERS AND RIBS.
Effect of Flexual Rigidity of Ribs on Buckling Load.
(Stringers and ribs not twisted - exact solution)
I_R = moment of inertia of section of ribs
I_S = " " " " stringers.
2b = spacing of ribs
2d = " " stringers
n-1 = number of ribs
m-1 = " " stringers
not counting edge members.
σ = " " half waves along stringers in buckled form.
Rigidity of ribs - Value of $\left\{\frac{I_R b^3}{I_S d^3}\right\}^{\frac{1}{2}} \frac{6(1-\cos\frac{\pi}{n})}{2+\cos\frac{\pi}{n}}$
Envelope for n → ∞
(any value of σ/n possible)
$\frac{\sigma}{n}=\frac{1}{6}$
$\frac{\sigma}{n}=\frac{2}{6}$
$\frac{\sigma}{n}=\frac{3}{6}$
$\frac{\sigma}{n}=\frac{4}{6}$
$\frac{\sigma}{n}=\frac{5}{6}$
0
1
2
3
4
5
10
20
30
40
50
60
70
80
90
100
Load - % of inter-rib buckling load.

Fig. 3.

THE BUCKLING OF GRIDS OF STRINGERS AND RIBS.

Effect of Flexural Rigidity of Ribs on Buckling Load.

(Stringers and ribs not twisted – approximate solution)

I_R = moment of inertia of section of ribs.
I_S = " " " " stringers.
$2b$ = spacing of ribs.
$2d$ = " " stringers.
$n-1$ = number of ribs
$m-1$ = " " stringers } not counting edge members.
σ = " " half waves along stringers in buckled form.

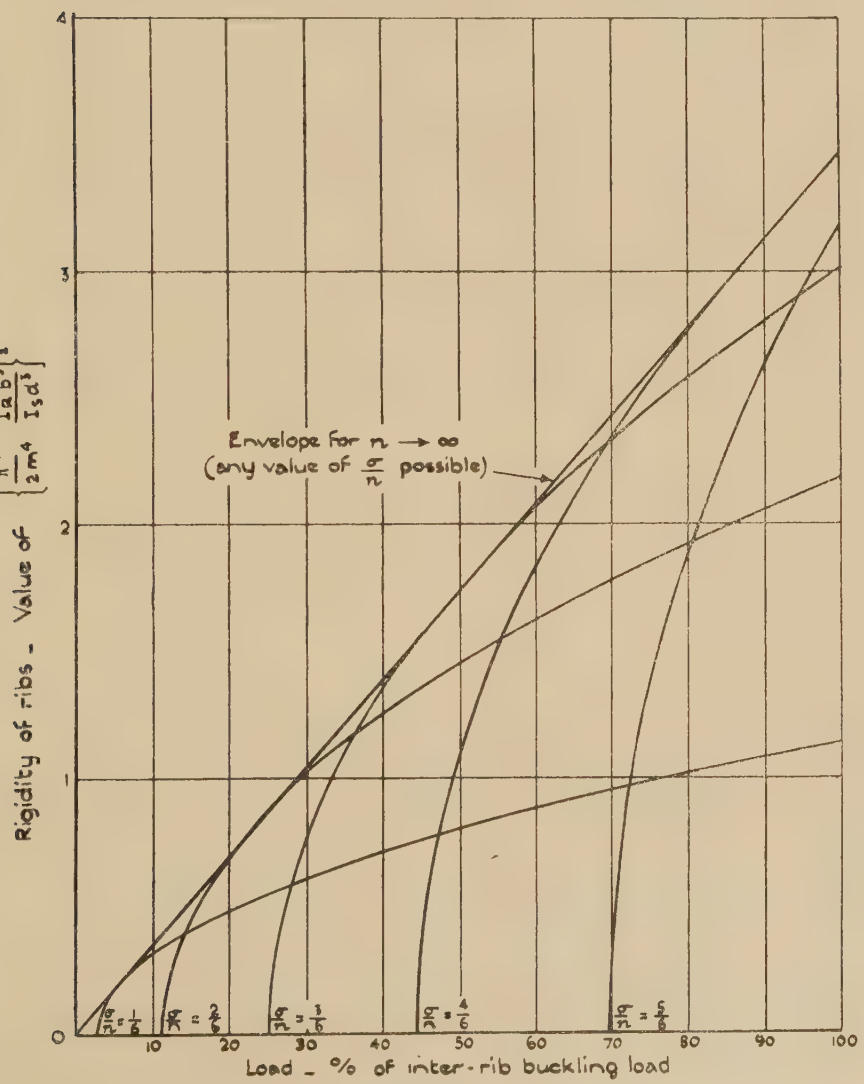


Fig. 4.

diagram, a straight line without the upward bend as θ approaches π which appears in Fig. 3, while the individual curves for definite values of σ/n are now true parabolas which would be represented by straight lines on a diagram similar to Fig. 2.

The failure of the approximate solution as θ approaches π is due to the considerable distortion of the ribs and stringers from the simple sine-wave form at high values of H/G. At low values of H/G, the loads imposed on the stringers by the ribs are relatively small and the over-riding importance of the end load tends to preserve the sine-wave form of the buckled strut, but when the value of H/G is sufficiently great to cause θ to approach π, the heavier lateral loads tend to emphasise the cubic form appropriate to a laterally loaded beam.

Acknowledgments. This work has formed part of the research programme of the Aeronautical Research Committee carried out at the National Physical Laboratory, and the authors are indebted to the Committee for permission to publish the work.

National Physical Laboratory,
Teddington, Middlesex.

THE BUCKLING OF A THIN SHEET TRANSVERSELY STIFFENED

By H. L. COX *and* H. E. SMITH.

[Received 1 June, 1938.—Read 16 June, 1938.—Revised 23 February, 1939.]

Synopsis.

The paper describes a theoretical analysis of the stability of a thin rectangular sheet with transverse ribs, simply supported at all edges, when subjected to compressive end load. The method is the Rayleigh-Ritz energy method. The results obtained in the previous paper by the same authors, "The buckling of grids of stringers and ribs", have been freely used. A "buckling equation" has been obtained from which it is possible to calculate the requisite flexural stiffness of the ribs in any case, assuming a value for their torsional rigidity.

Specification of problem and statement of assumptions.

1. A plane rectangular sheet of thickness $2h$, of total length $2B$ and total width $2D$, is divided into n rectangles, each of length $2b\ (=2B/n)$ by $n-1$ ribs. (See Fig. 2*a*.)

It will be assumed:

2. That the construction is subjected to a uniform (average) overall compressive strain e in the direction of the length (so that the edges parallel to the ribs remain parallel but approach each other by the distance $2Be$).

3. That the edges of the sheet and the ends of all the ribs are held in a plane but are not otherwise constrained (simple support).

4. That the neutral axes of the sections of all the ribs lie in the plane of the sheet.

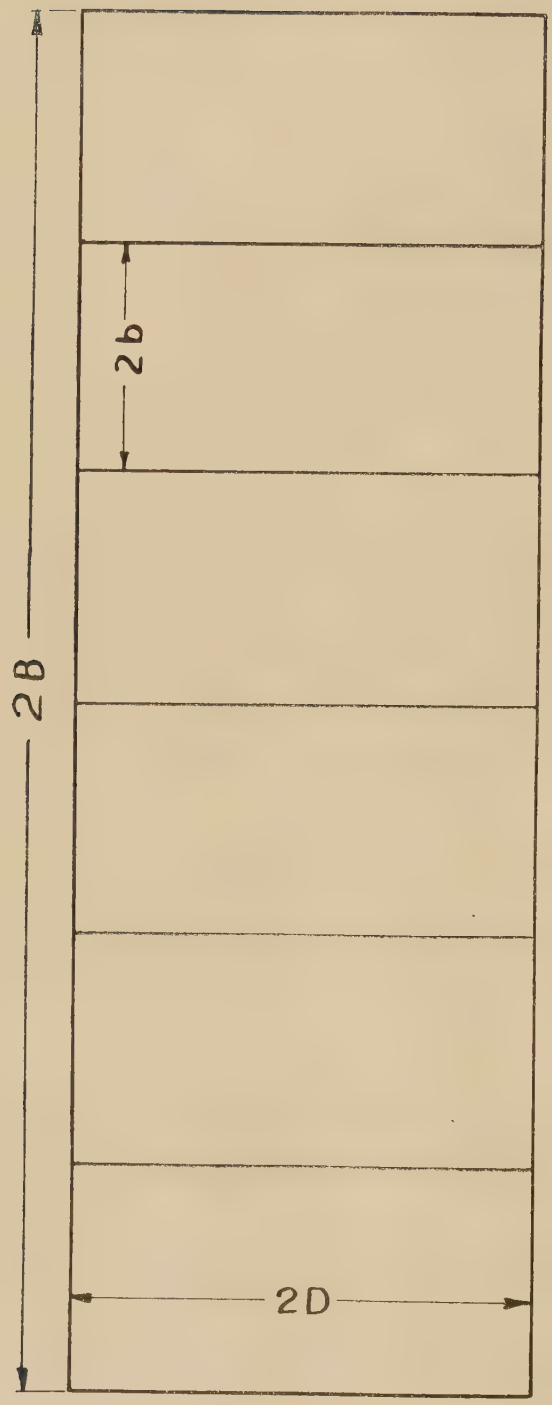


Fig. 2*a*.
Transversely stiffened sheet.
($n = 6$).

Fig. 2*b*.
Mid-section of buckled sheet.
($n = 6$, $s = 5$.)

5. That all the ribs have the same flexural rigidity EI and the same torsional rigidity CJ, except the end ribs, which have a torsional rigidity $\frac{1}{2}CJ$.

6. That the buckled form of the sheet is given by

$$w = \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \delta_{rs} \sin \frac{r\pi x}{2D} \sin \frac{s\pi y}{2B}, \tag{1}$$

where w is the deflection perpendicular to the sheet at a point whose coordinates are x, y, with regard to the width and length of the sheet respectively.

Since this expression is capable of representing any buckled form, and since no restriction of generality of the expression is later introduced, the result obtained is exact. This point and the physical interpretation of the results, particularly with regard to the various possible buckled forms, are dealt with at greater length in the previous paper.

The total deformation energy of the system is made up of

(1) Flexural energy of deformation of the sheet, expressed by

$$\frac{Eh^3}{3(1-\sigma^2)} \int_0^{2B} \int_0^{2D} (\nabla^2 w)^2 \, dy \, dx.$$

This is equal to

$$\frac{Eh^3}{3(1-\sigma^2)} \cdot \frac{\pi^4}{16BD} \Sigma \left(r^2 \frac{B}{D} + s^2 \frac{D}{B} \right)^2 \delta_{rs}^2.$$

(2) Flexural energy of the ribs. It is shown in the previous paper that this energy is equal to

$$H \sum_r \Sigma'_s r^4 (\Delta)^2,$$

where

$$H = EI \frac{\pi^4 n}{64D^3}, \qquad \Delta = \delta_{r,s} - \delta_{r,2n-s} + \delta_{r,2n+s} - \cdots,$$

and Σ'_s denotes summation for all values of s less than n.

(3) Torsional energy of the ribs. From Part II of the previous paper, this energy is equal to

$$K \left[\sum_r \Sigma'_s r^2 (\Delta')^2 + \tfrac{1}{2} \sum_r r^2 ({}_n\Delta_r')^2 + \tfrac{1}{2} \sum_r r^2 ({}_0\Delta_r')^2 \right],$$

where $$K = CJ\frac{\pi^4 n}{64B^2D},$$

$$\Delta' = s\delta_{r,s} + (2n-s)\,\delta_{r,2n-s} + (2n+s)\,\delta_{r,2n+s} + \ldots,$$

$${}_n\Delta_r' = 2(n\delta_{r,n} + 3n\delta_{r,3n} + \ldots), \qquad {}_0\Delta_r' = 2(2n\delta_{r,2n} + 4n\delta_{r,4n} + \ldots).$$

(4) Direct stress energy of the sheet. This is equal to

$$4EBDh\left\{e^2 - \Sigma\frac{es^2\pi^2}{16B^2}\delta_{r,s}^2\right\}.$$

(omitting terms in δ^4).

The total energy of deformation is thus

$$\frac{EDh\pi^2}{16B}e_s\Sigma\left(r^2\frac{B}{D} + s^2\frac{D}{B}\right)^2\delta_{r,s}^2 + H\sum_r\sum_s{}'r^4(\Delta)^2$$

$$+K\left[\sum_r\sum_s{}'r^2(\Delta')^2 + \tfrac{1}{2}\sum_r r^2({}_n\Delta_r')^2 + \tfrac{1}{2}\sum_r r^2({}_0\Delta_r')^2\right]$$

$$+4EBDh\left\{e^2 - \Sigma\frac{es^2\pi^2}{16B^2}\delta_{r,s}^2\right\}, \qquad (2)$$

where $e_s = \dfrac{\pi^2h^2}{3(1-\sigma^2)D^2}$ is the buckling strain of the unstiffened sheet (assuming that $B > D$).

It is seen from examination of this expression that, with respect to r, $r = 1$ gives a minimum value. Thus higher values of r need not be considered, and we may put $r = 1$ in all future working. We may now drop the first suffix of the parameters, and at the same time for the reasons stated in the previous paper, we write the other suffix in the form $2Nn \pm s$, where N is any positive integer and s has the range of values from 1 to $n-1$. The form $2Nn \pm s$ is also intended to include the value s.

Henceforth s will be treated as a constant, its several values 1, 2, 3, ..., $n-1$ being considered individually.

Differentiating the expression for total energy with respect to $\delta_{2Nn\pm s}$ and equating to zero, we get

$$\left[EBDh\frac{\pi^2}{16B^2}e_s\left\{\frac{B}{D} + \frac{(2Nn\pm s)^2D}{B}\right\} - 4EBDh\frac{\pi^2}{16B^2}e(2Nn\pm s)^2\right]\delta_{2Nn\pm s}$$

$$\pm H\Delta + K(2Nn\pm s)\Delta' = 0. \qquad (3)$$

Dividing throughout by $EBDh\frac{D^2}{B^2}\frac{\pi^2}{16B^2}e_s$, and writing

$$EI' = \frac{EIn}{2B} = \text{flexural rigidity of ribs per unit length of sheet,}$$

$$CJ' = \frac{CJn}{2B} = \text{torsional rigidity of ribs per unit length of sheet,}$$

$$EI_s' = \tfrac{2}{3}\frac{Eh^3}{1-\sigma^2} = \text{flexural rigidity of sheet per unit width,}$$

we obtain

$$\left[\left\{(2Nn\pm s)^2+\frac{B^2}{D^2}\right\}^2-4\frac{B^2}{D^2}\frac{e}{e_s}(2Nn\pm s)^2\right]\delta_{2Nn\pm s}$$
$$\pm\frac{EI'}{EI_s'}\cdot\frac{B^4}{D^4}\Delta+\frac{CJ'}{EI_s'}\cdot\frac{B^2}{D^2}(2Nn\pm s)\Delta'=0. \quad (4)$$

Then, writing

$$\frac{B}{D}=\rho, \qquad \frac{EI'}{EI_s'}=X, \qquad \frac{CJ'}{EI_s'}=Y,$$

we have

$$\left[\{(2Nn\pm s)^2+\rho^2\}^2-4\rho^2\frac{e}{e_s}(2Nn\pm s)^2\right]\delta_{2Nn\pm s}$$
$$\pm X\rho^4\Delta+Y\rho^2(2Nn\pm s)\Delta'=0. \quad (5)$$

Dividing through in (5) by the coefficient of $\delta_{2Nn\pm s}$, taking successive values of N, and summing in the appropriate manner we get the two equations

$$\Delta(1+X\rho^4\Sigma_0)+Y\rho^2\Sigma_1\Delta'=0, \quad (6)$$

$$\Delta'(1+Y\rho^2\Sigma_2)+X\rho^4\Sigma_1\Delta=0, \quad (7)$$

where

$$\Sigma_0=\sum_0^\infty\frac{1}{\{(2Nn\pm s)^2+\rho^2\}^2-4\rho^2(2Nn\pm s)^2e/e_s},$$

$$\Sigma_1=\sum_0^\infty\frac{\pm(2Nn\pm s)}{\{(2Nn\pm s)^2+\rho^2\}^2-4\rho^2(2Nn\pm s)^2e/e_s},$$

$$\Sigma_2=\sum_0^\infty\frac{(2Nn\pm s)^2}{\{(2Nn\pm s)^2+\rho^2\}-4\rho^2(2Nn\pm s)^2e/e_s}.$$

The sums of these series are given below.

Eliminating Δ and Δ' from equations (6) and (7), we get as the buckling equation

$$XY\rho^6(\Sigma_1)^2-(1+X\rho^4\Sigma_0)(1+Y\rho^2\Sigma_2)=0, \tag{8}$$

where

$$X=\frac{EI'}{EI_s'}=\frac{\pi^2 In}{4BD^2h}e_s, \qquad Y=\frac{CJ'}{EI_s'}=\frac{\pi^2 CJn}{4EBD^2h}e_s, \qquad \rho=\frac{B}{D},$$

$$\Sigma_0=\frac{\pi}{2n\{(\rho\epsilon)^2-(\rho/\epsilon)^2\}}\left[\frac{\sin(\epsilon\theta)/\rho\epsilon}{\cos\epsilon\theta-\cos(s\theta/\rho)}-\frac{\epsilon\sin(\theta/\epsilon)/\rho}{\cos(\theta/\epsilon)-\cos(s\theta/\rho)}\right],$$

$$\Sigma_1=\frac{\pi}{2n\{(\rho\epsilon)^2-(\rho/\epsilon)^2\}}\cdot\frac{\sin(s\theta/\rho)}{1-\cos(s\theta/\rho)} \times\left[\frac{1-\cos(\epsilon\theta)}{\cos\epsilon\theta-\cos(s\theta/\rho)}-\frac{1-\cos(\theta/\epsilon)}{\cos(\theta/\epsilon)-\cos(s\theta/\rho)}\right],$$

$$\Sigma_2=\frac{\pi}{2n\{(\rho\epsilon)^2-(\rho/\epsilon)^2\}}\left[\frac{\rho\epsilon\sin(\epsilon\theta)}{\cos(\epsilon\theta)-\cos(s\theta/\rho)}-\frac{\rho\sin(\theta/\epsilon)/\epsilon}{\cos(\theta/\epsilon)-\cos(s\theta/\rho)}\right],$$

and where

$$\epsilon=\sqrt{\left(\frac{e}{e_s}\right)}+\sqrt{\left(\frac{e}{e_s}-1\right)}, \qquad e_s=\frac{\pi^2h^2}{3(1-\sigma^2)D^2}, \qquad \theta=\frac{\pi\rho}{n}.$$

Interpretation of results.

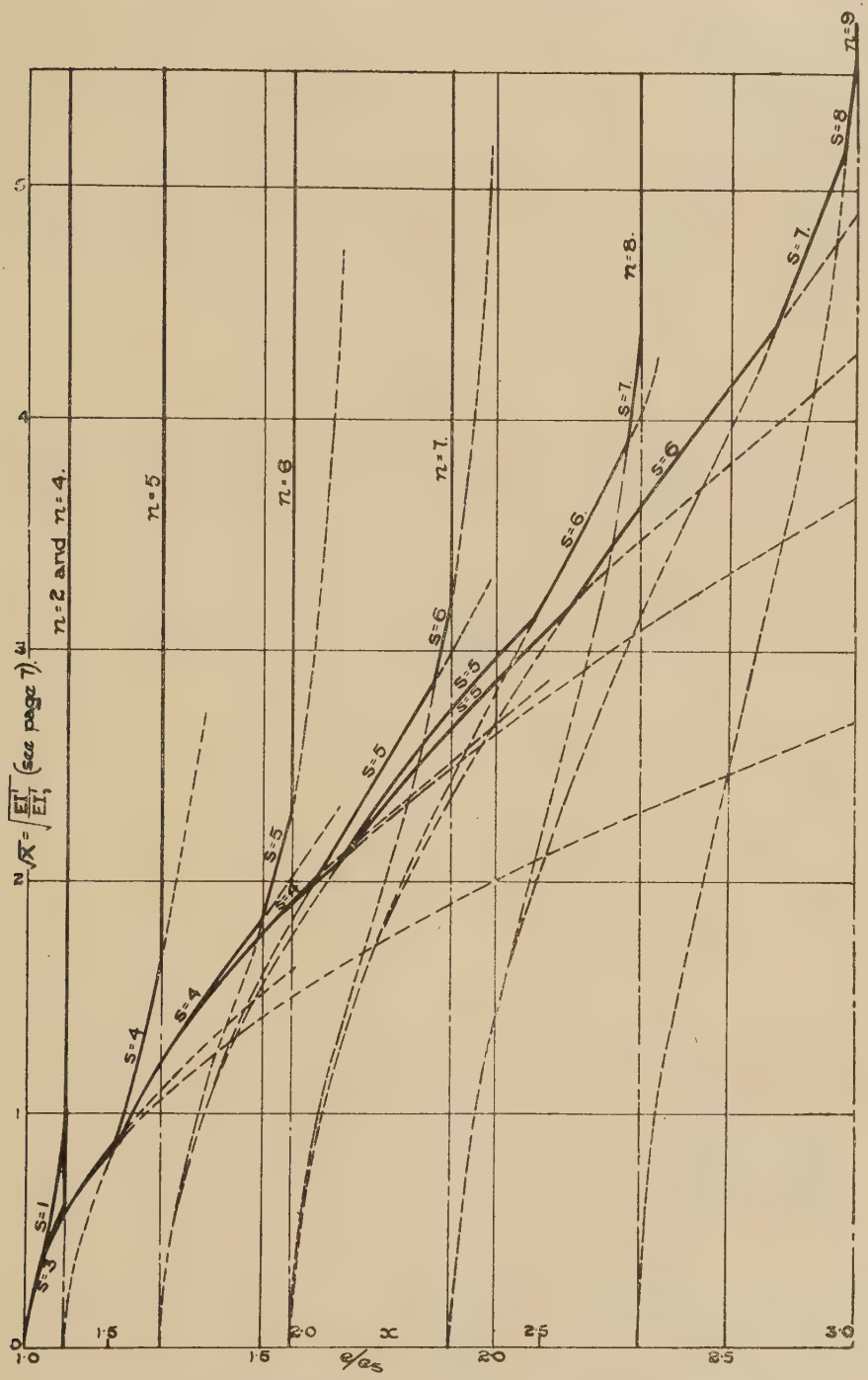
For a sheet of given dimensions and with ribs of specified *torsional* rigidity, it is possible, with the aid of equation (8), to determine the requisite *flexural* rigidity of the ribs to stiffen the structure for a given value of e and for each possible value of s. Thus the type of buckling, the buckling load, and the buckling strain are determined. It is, however, not so simple to interpret equation (8) in general terms. In order to simplify the problem we shall therefore neglect the torsional rigidity of the ribs. Equation (8) (or equation (6)) then reduces to

$$1+X\rho^4\Sigma_0=0.$$

When we insert the value of Σ_0, this becomes

$$X\left\{\frac{\epsilon\sin(\theta/\epsilon)}{\cos(\theta/\epsilon)-\cos(s\theta/\rho)}-\frac{\sin(\epsilon\theta)/\epsilon}{\cos(\epsilon\theta)-\cos(s\theta/\rho)}\right\}=\frac{2}{\theta}\left(\epsilon^2-\frac{1}{\epsilon^2}\right). \tag{9}$$

It is now possible to calculate values of X for ranges of values of ϵ (or e/e_s), ρ and s. Fig. 1 shows the results of such calculations for the



Buckling of a Thin Sheet Transversely Stiffened.
Amount of Stiffening required in relation to Buckling Strain.

Fig. 1.

cases $\rho = 3$, all values of n up to 9 and all values of s. It will be seen that, as in the case of the grid of stringers and ribs without sheet, the set of curves for one value of n consists of overlapping curves, the curve for each value of s rising above that for the previous value. In Fig. 1 the highest value of X for each value of e/e_s is shown by a thick line. This line becomes vertical at the inter-rib buckling value of e/e_s. It should be noted that very considerable extra rib stiffening is required against $n-1$ buckling as compared with $n-2$ buckling. Fig. 2b shows the form of a buckled sheet for which $n = 6$ and $s = 5$.

Acknowledgments.

This work has formed part of the research programme of the Aeronautical Research Committee carried out at the National Physical Laboratory, and the authors are indebted to the Committee for permission to publish the work.

National Physical Laboratory,
Teddington, Middlesex.

A UNIQUENESS THEOREM FOR MEAN VALUES OF ANALYTIC FUNCTIONS

By W. H. J. FUCHS.

[Received and read 17 November, 1938.]

1. The purpose of this paper is to prove the following theorem which was conjectured by A. E. Ingham.

THEOREM 1. *If $f(z)$ and $g(z)$ are regular in the annulus $A: a<|z|<b$ and if, for a fixed positive λ, $M_\lambda(r, f) = M_\lambda(r, g)$ for an infinity of r with a limiting value R $(a<R<b)$,*

where
$$M_\lambda(r, f) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(re^{i\theta})|^\lambda d\theta\right)^{1/\lambda}$$

and
$$M_\lambda(r, g) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |g(re^{i\theta})|^\lambda d\theta\right)^{1/\lambda},$$

*then**
$$M_\lambda(r, f) \equiv M_\lambda(r, g),$$

in A.

2. *Proof in a special case.* The proof is very simple when $f(z)$ and $g(z)$ have no zeros in A. In this case $F(z) = \big(f(z)\big)^{\frac{1}{2}\lambda}$ and $G(z) = \big(g(z)\big)^{\frac{1}{2}\lambda}$ are regular† at every point of A and their absolute values are single-valued. It follows that $F(z)$ and $G(z)$ are either single-valued or that the branches of $F(z)$ [or of $G(z)$] differ by factors of the form $e^{\alpha i}$ only. Suppose that

* We use the symbol $\equiv$ as an abbreviation of "is identically equal to."

† By "$h(z)$ is regular at the point $z = z_0$" we mean that $h(z)$ can be expanded into a convergent Taylor series $\sum_{n=0}^{\infty} c_n(z-z_0)^n$ for sufficiently small $|z-z_0|$. By "$h(z)$ is regular in the domain D" we mean that $h(z)$ is regular at every point of D and single-valued in D.

the result of continuation along a complete circle $|z| = \text{const.}$ in the positive sense is to multiply $F(z)$ by the factor $e^{2\nu\pi i}$. Then

$$F(z) = z^{\nu} \sum_{n=-\infty}^{\infty} a_n z^n \quad (a < |z| < b),$$

since $z^{-\nu} F(z)$ is regular in A (Monodromiesatz, Laurent's theorem). Hence, by Parseval's theorem,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(re^{i\theta})|^{\lambda} d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} |F(re^{i\theta})|^2 d\theta = r^{2\nu} \sum_{n=-\infty}^{\infty} |a_n|^2 r^{2n} \quad (a < r < b) \tag{2.1}$$

and similarly

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |g(re^{i\theta})|^{\lambda} d\theta = r^{2\mu} \sum_{n=-\infty}^{\infty} |\beta_n|^2 r^{2n} \quad (a < r < b).$$

That is to say, the mean values are analytic functions of the *complex* variable r in $a < |r| < b$. Therefore, by the well-known uniqueness theorem for analytic functions, they are identical under the hypotheses of Theorem 1.

For this proof it is sufficient to assume that $[f(z)]^{\frac{1}{2}\lambda}$ and $[g(z]^{\frac{1}{2}\lambda}$ are regular at every point of A, a slightly weaker condition than that they should have no zeros in A.

3.1. *The general case.* To simplify the appearance of our formulae a little we prove the slightly more general

THEOREM 2. *If* (i)

$$\left.\begin{aligned} f(z) &= z^{\rho} \prod_{\nu=1}^{M} (\omega_\nu - z)^{a_\nu} \sum_{n=-\infty}^{\infty} a_n z^n \qquad (0 < a_\nu < 1, \ \ 0 \leqslant \rho < 1) \\ \text{and} \qquad & \\ g(z) &= z^{\rho'} \prod_{\nu=1}^{M'} (\omega_\nu' - z)^{a_\nu'} \sum_{n=-\infty}^{\infty} a_n' z^n \ \ (0 < a_\nu' < 1, \ \ 0 \leqslant \rho' < 1), \end{aligned}\right\} \tag{3.11}$$

where the two infinite series are convergent in $a < |z| < b$, *and if* (ii)

$$M_2(r, f) = M_2(r, g)$$

for an infinity of r *with a limiting value* R $(a < R < b)$, *where*

$$M_2{}^2(r, f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(re^{i\theta})|^2 \, d\theta$$

and
$$M_2^2(r, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |g(re^{i\theta})|^2 d\theta,$$

then
$$M_2(r, f) \equiv M_2(r, g).$$

The proof is based on the same ideas as above, but the mean values are now no longer given by *one* analytic function for all values of $r(a < r < b)$. The circles $|z| = |\omega_\nu|$ and $|z| = |\omega_\nu'|$ subdivide A into a finite number of annuli in each of which $f(z)$ and $g(z)$ are regular at every point. Therefore, by § 2, $M_2^2(r, f)$ and $M_2^2(r, g)$ are analytic functions in each of these annuli. Our proof of Theorem 2 falls into two parts. We show:

(1) that the hypotheses ensure the equality of the analytic functions representing $M_2^2(r, f)$ and $M_2^2(r, g)$ in one of the sub-annuli, and

(2) that the function representing $M_2^2(r, f)$ in any one annulus completely determines the functions representing $M_2^2(r, f)$ in the other annuli.

If R differs from all the numbers $|\omega_\nu|$ and $|\omega_\nu'|$, the proof of (1) is already included in § 2. In the case when R is equal to one of these numbers we require the following

LEMMA 1. *Let $P(x)$ have the following convergent expansion for positive $x < H$:*
$$P(x) = \sum_{h=1}^{\infty} A_h x^{m_h} + \log x \sum_{h=1}^{\infty} B_h x^{n_h},$$

where the logarithm is real, $0 \leqslant m_1 < m_2 \ldots$, $0 < n_1 < n_2 \ldots$ and m_h and n_h tend to infinity with h. Then $P(x)$ is identically zero, if $x = 0$ is a limit point of zeros of $P(x)$.

Suppose the first non-vanishing coefficients in the respective series to be A_j and B_l. Then, for small x,

$$P(x) = A_j x^{m_j} + B_l \log x . x^{n_l} + x^{m_{j+1}} \sum_{h=j+1}^{\infty} A_h x^{m_h - m_{j+1}} + \log x . x^{n_{l+1}} \sum_{h=l+1}^{\infty} B_h x^{n_h - n_{l+1}}$$

$$= \begin{cases} A_j x^{m_j}\big(1 + o(1)\big) & (m_j < n_l), \\ B_l x^{n_l} \log x \big(1 + o(1)\big) & (m_j \geqslant n_l), \end{cases}$$

as x tends to 0. Hence $P(x)$ cannot have zeros arbitrarily near to the origin. This is a contradiction.

To prove (2) it is sufficient to show that from the mean value $M_2{}^2(r, f)$ for values of r just less than 1, we can determine $M_2{}^2(r, f)$ uniquely for values of r just greater than 1, since the general case can be reduced to this by the substitutions $z' = Kz$ or $z' = K/z$. We can find two constants c and d $(c < 1 < d)$ such that for $c < |z| < d$ we may write

$$f(z) = z^\gamma \prod_{\nu=1}^{N} (\omega_\nu - z)^{\alpha_\nu} \phi(z) \quad (0 \leqslant \gamma < 1,\ 0 < \alpha_\nu < 1,\ |\omega_\nu| = 1)$$

$$= z^\gamma h(z), \tag{3.12}$$

where $\phi(z)$ is regular in $c < |z| < d$. The equation (3.12) is obtained from (3.11) by expanding the terms $(\omega - z)^\alpha$ with $|\omega| > 1$ into a power series in z and those with $|\omega| < 1$ into a power series in $1/z$. Once the determination of the α-th powers has been fixed, $h(z)$ is regular in $c < |z| < 1$. We can also write

$$f(z) = z^{\gamma + \sum_{\nu=1}^{N} \alpha_\nu} \prod_{\nu=1}^{N} (\omega_\nu z^{-1} - 1)^{\alpha_\nu} \phi(z)$$

$$= z^{\gamma + \sum_{\nu=1}^{N} \alpha_\nu} k(z), \tag{3.13}$$

where $k(z)$ is regular in $1 < |z| < d$.

There are, of course, analogous expressions for $g(z)$. We use the same letters for corresponding expressions connected with $f(z)$ and $g(z)$ and distinguish those relating to $g(z)$ by a dash. Thus the zeros of $g(z)$ on $|z| = 1$ are ω_ν' $(\nu = 1, 2, \ldots, N')$.

3.2. *Expression of $M_2{}^2(r, f)$ as a contour-integral.* For $c < r < 1$,

$$\int_{-\pi}^{\pi} |f(re^{i\theta})|^2 d\theta = r^{2\gamma} \int_{-\pi}^{\pi} h(re^{i\theta}) \overline{h(re^{i\theta})}\, d\theta$$

$e^{i\theta} = \omega$:

$$= -ir^{2\gamma} \oint_{|\omega|=1} h(r\omega) \overline{h(\bar{r}/\bar{\omega})}\, \omega^{-1} d\omega$$

$\omega = r\zeta$:

$$= -ir^{2\gamma} \oint_{|\zeta|=1/r} h(r^2\zeta) \bar{h}(1/\zeta) \frac{d\zeta}{\zeta}, \tag{3.21}$$

where $\bar{h}(u)$ stands for the "conjugate function" $\overline{h(\bar{u})}$; this is a regular function, if $h(u)$ is regular.

The only singularities of $h(z)$ in $c<|z|<d$ are the points $z=\omega_\nu$. $h(z)$ is regular in $c<|z|<d$ after cuts have been made from the points ω_ν to infinity. We regard these cuts as straight lines going radially outwards, but their shape may be varied within certain limits. Hence $h(r^2\zeta)$ is a regular function of ζ in the annulus

$$c<|r^2\zeta|<d \tag{3.22}$$

with cuts from the points $\zeta=\omega_\nu r^{-2}$ to infinity. $\bar{h}(\zeta^{-1})$ is regular in the annulus

$$c<|\zeta^{-1}|<d \tag{3.23}$$

with cuts from $\bar{\zeta}^{-1}=\omega_\nu$ to $\bar{\zeta}^{-1}=\infty$, *i.e.* from $\zeta=\omega_\nu$ to $\zeta=0$. The integrand of (3.21) *qua* function of ζ is therefore regular in the common part of (3.22) and (3.23), *i.e.* in

$$\max\left(\frac{1}{d}, \frac{c}{|r^2|}\right)<|\zeta|<\min\left(\frac{1}{c}, \frac{d}{|r^2|}\right) \tag{3.24}$$

after the cuts indicated have been made. The path of integration may therefore be replaced by any other path l inside this domain encircling the origin once in the positive sense. We put $r^2=u$ and study the integral (3.21) as a function of u. The variable u may now assume complex values. The function

$$F(u)=-iu^\nu\oint_l h(u\zeta)\,\bar{h}(1/\zeta)\,\frac{d\zeta}{\zeta} \tag{3.25}$$

is an analytic function of u and the representation by the integral (3.25) is valid for all values u_1 of u which can be connected to a real $u=u_0$ $(c^2<u_0<1)$ by a continuous path $u=u(t)$ $[0\leqslant t\leqslant 1;\ u_0=u(0),\ u_1=u(1)]$ satisfying the following conditions:

(i) corresponding to every $u=u(t)$ the domain (3.24) exists, *i.e.*

$$\max\left(\frac{1}{d}, \frac{c}{|u|}\right)<\min\left(\frac{1}{c}, \frac{d}{|u|}\right);$$

(ii) corresponding to every u there is a path $l'=l(t)$ satisfying all the conditions imposed on l. As t varies from 0 to 1 the various $l(t)$ can be obtained from $l(0)$ by successive continuous deformations without crossing any cut. It is permissible to alter the shape of the cuts continuously with t so as to satisfy the last requirement.

In particular we can continue along the following two paths:

(*a*) from a real u ($c^2 < u < 1$) through the half-plane $\mathfrak{J}(u) < 0$ to a real $u > 1$ and sufficiently close to 1, keeping all the time close to the real axis [result of continuation $F^+(u)$];

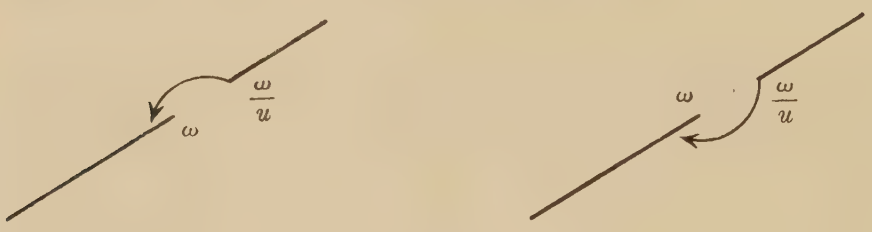
(*b*) the same through the half-plane $\mathfrak{J}(u) > 0$. [Result $F^-(u)$.]

u-plane. *u-plane.*

(*a*) (*b*)

Motion of cuts with varying u:

(*a*) (*b*)



Path of integration after continuation:

(*a*) (*b*)

It is not difficult to see that $h(u\zeta)\overline{h}(1/\zeta)$ is real on the part $|\zeta| = 1/r$ of the new path of integration. This follows either from a consideration of the change of amplitude of $h(u\zeta)\overline{h}(1/\zeta)$ when u moves along (*a*) or (*b*) and ζ moves at the same time in accordance with the equation $r\zeta = e^{i\gamma}$, when γ is suitably chosen so that ζ does not collide with a cut, or from a direct consideration of the expression (3.12). The contribution of the

circle is therefore

$$-ir^{2\gamma}\int_{|\zeta|=1/r} h(r^2\zeta)\bar{h}(1/\zeta)\,d\zeta/\zeta = r^{2\gamma}\int_{-\pi}^{\pi} h(re^{i\theta})\overline{h(re^{i\theta})}\,d\theta$$

$$= 2\pi M_2{}^2(r, f) \tag{3.31}$$

Since the singularities ω_ν and ω_ν/u are branch points of the integrand at which it behaves in the way that v^{α_ν} behaves at the origin, the values of the integrand on opposite sides of a cut differ merely by a factor $e^{2\pi\alpha_\nu i}$. The effect of encircling a singularity in the positive sense is to multiply the integrand by this factor. Hence

$$\int_A^B + \int_B^C + \int_C^D = (1-e^{2\pi\alpha_\nu i})\int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\,\frac{d\zeta}{\zeta}, \tag{3.32}$$

where the integral is taken from B to C. By (3.31) and (3.32),

$$F^-(u) = 2\pi M_2{}^2(r, f) - i\sum_{\nu=1}^{N}(1-e^{2\pi\alpha_\nu i})u^\gamma\int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\,\frac{d\zeta}{\zeta}. \tag{3.33}$$

And, since on $A'B'$ the integrand is equal to $e^{2\pi\alpha i}h(u\zeta)\bar{h}(1/\zeta)$, on $B'C'$ to $e^{4\pi\alpha i}h(u\zeta)\bar{h}(1/\zeta)$, and on $C'D'$ to $e^{2\pi\alpha i}h(u\zeta)\bar{h}(1/\zeta)$, with $h(u\zeta)\bar{h}(1/\zeta)$ determined as in (3.32), we have

$$F^+(u) = 2\pi M_2{}^2(r, f) - i\sum_{\nu=1}^{N} e^{2\pi\alpha_\nu i}(1-e^{2\pi\alpha_\nu i})u^\gamma\int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\,\frac{d\zeta}{\zeta}. \tag{3.34}$$

3.4. *The nature of the singularity* $u = 1$. We next study the function of u

$$\int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\,d\zeta/\zeta.$$

Substituting from (3.12) and defining a new variable t by the formula

$$u\zeta = \omega_\nu[1+t(u-1)],$$

we see that, for u near 1 and ζ near ω_ν,

$$h(u\zeta)\bar{h}(1/\zeta) = [t(1-u)^{\alpha_\nu}][1+t(u-1)-u]^{\alpha_\nu}H(t, u)$$

$$= (1-u)^{2\alpha_\nu}t^{\alpha_\nu}(1-t)^{\alpha_\nu}H(t, u),$$

where $H(t, u)$ is a regular function of u for every real t $(0 \leqslant t \leqslant 1)$ when

u is sufficiently near to 1. And

$$\int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\,d\zeta/\zeta = (1-u)^{2a_\nu+1}\int_0^1 G(t,u)\,dt,$$

where the integral is a regular function of u for small $|u-1|$. Hence

$$u^\gamma \int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\,d\zeta/\zeta = (1-u)^{2a_\nu+1}\phi_\nu(u-1),$$

where $\phi_\nu(u-1)$ is regular in $|u-1|<l$ for some $l>0$. If we take a real value of $u<1$, $\arg(1-u)=0$, and choose the determinations of h and $\bar{h}$ so that for real values of $x<1$, $h(x)$ and $\bar{h}(x)$ are conjugate complex numbers, then the substitution $u\zeta=\bar{\eta}^{-1}$ shows that

$$\int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\frac{d\zeta}{\zeta} = \int_{\omega_\nu/u}^{\omega_\nu} h(1/\bar{\zeta})\bar{h}(u\zeta)\frac{d\zeta}{\zeta},$$

i.e. $\phi_\nu(u-1)=\overline{\phi_\nu(u-1)}$, and so $\phi_\nu(u-1)$ is real.

If we start from a $u<1$ with these determinations of the h and continue to $u>1$ along the path (b) of the previous section, the integral is continued into the integral in (3.32), $\arg(1-u)$ decreases to $-\pi$ as u varies along (b). Hence we see that the integral in (3.32) is equal to

$$u^{-\gamma}(1-u)^{2a_\nu+1}\phi_\nu(u-1) \quad [\arg(1-u)=-\pi]. \tag{3.41}$$

Subtracting (3.33) from (3.34), we have

$$F^+(u)-F^-(u) = i\sum_{\nu=1}^{N}(1-e^{2\pi a_\nu i})^2(1-u)^{2a_\nu+1}\phi_\nu(u-1) \quad [\arg(1-u)=-\pi]. \tag{3.42}$$

This equation asserts that $M_2{}^2(r,f)$ can be analytically continued around the point $r=1$ from values of $r<1$ to real values of $r>1$. The result of continuation through the upper half-plane is different from that of continuation through the lower half-plane. The equation (3.42) gives an expression for the difference of the two results. By the theory of analytic continuation we may now drop the condition $\arg(1-u)=-\pi$. The interpretation of (3.42) is now: The function $F(u)$ $[=M_2{}^2(r,f)$ for $r<1]$ can be continued analytically round the point $u=1$. The result of continuing once round this point in the positive sense is to add the right-hand side of (3.42) to the initial value.

Consider now the function

$$\Xi(u-1) = F(u) - i \sum_{a_\nu \neq \frac{1}{2}} \frac{(1-e^{2\pi a_\nu i})^2}{e^{4\pi a_\nu i}-1} (1-u)^{2a_\nu+1} \phi_\nu(u-1)$$

$$- \frac{2}{\pi} \log(1-u) \sum_{a_\nu = \frac{1}{2}} (1-u)^2 \phi_\nu(u-1). \quad (3.43)$$

$\Xi(u-1)$ is regular at every point of the annulus $0 < |u-1| < l$. When $\arg(1-u)$ is increased by 2π,

$$\Xi[e^{2\pi i}(u-1)] = F(u) + i \sum_{\nu=1}^{N} (1-e^{2\pi a_\nu i})^2 (1-u)^{2a_\nu+1} \phi_\nu(u-1)$$

$$- i \sum_{a_\nu \neq \frac{1}{2}} \frac{(1-e^{2\pi a_\nu i})^2}{e^{4\pi a_\nu i}-1} e^{4\pi a_\nu i} (1-u)^{2a_\nu+1} \phi_\nu(u-1)$$

$$- \frac{2}{\pi} \log(1-u) \sum_{a_\nu = \frac{1}{2}} (1-u)^2 \phi_\nu(u-1)$$

$$- i \sum_{a_\nu = \frac{1}{2}} 4(1-u)^2 \phi_\nu(u-1)$$

$$= \Xi(u-1).$$

Hence $\Xi(u-1)$ is regular in $0 < |u-1| < l$ and is bounded at $u-1 = 0$. By a well-known theorem $\Xi(u-1)$ is consequently regular in $|u-1| < l$. We write

$$\left.\begin{aligned} \sum_{a_\nu = a} \phi_\nu(u-1) &= E_a(u-1), \\ \sum_{a_\nu = a+\frac{1}{2}} (1-u)\phi_\nu(u-1) &= H_a(u-1) \quad (0 < a < \tfrac{1}{2}), \\ \sum_{a_\nu = \frac{1}{2}} (1-u)^2 \phi_\nu(u-1) &= \psi(u-1). \end{aligned}\right\} \quad (3.44)$$

(3.43) can be simplified a little, because

$$-i \frac{(1-e^{2ai})^2}{e^{4ai}-1}$$

is easily reduced to $\tan a$.

(3.43) now becomes

$$F(u) = \Xi(u-1) + \sum_a (1-u)^{2a+1} \{\cot \pi a \, H_a(u-1) - \tan \pi a \, E_a(u-1)\}$$

$$+ \frac{2}{\pi} \log(1-u) \psi(u-1). \quad (3.45)$$

The summation is over all a $(0 < a < \frac{1}{2})$ corresponding to which there exists an a_ν such that $a_\nu = a$ or $a = a + \frac{1}{2}$.

For real r $(c<r<1)$ and $u=r^2$, we have $F(u)=2\pi M_2^2(r,f)$. For real $r>1$, we have, from (3.33) and (3.41),

$$2\pi M_2^2(r,f)$$
$$=\Xi(u-1)-\sum_a e^{-2\pi a i}(u-1)^{2a+1}\{\cot\pi a\, H_a(u-1)-\tan\pi a\, E_a(u-1)\}$$
$$+\frac{2}{\pi}\log(u-1)\,\psi(u-1)-2i\psi(u-1)$$
$$-i\sum_a (u-1)^{2a+1}\{(1-e^{2\pi a i})\,e^{-2\pi a i}E_a(u-1)+(1+e^{2\pi a i})\,e^{-2\pi a i}H_a(u-1)\}$$
$$+2i\psi(u-1).$$

Since
$$e^{-2\pi a i}(\cot\pi a+i+ie^{2\pi a i})=\cot\pi a,$$
$$e^{-2\pi a i}(\tan\pi a-i+ie^{2\pi a i})=-\tan\pi a,$$

$$2\pi M_2^2(r,f)=\Xi(u-1)-\sum_a(u-1)^{2a+1}\{\cot\pi a\,H_a(u-1)+\tan\pi a\,E_a(u-1)\}$$
$$+\frac{2}{\pi}\log(u-1)\,\psi(u-1)\quad(r>1).\qquad(3.46)$$

There are, of course, analogous expressions for $M_2^2(r,g)$.

We can now completely prove the assertion (1) of § (3.1). Suppose, for example, that $M_2^2(r,f)=M_2^2(r,g)$ for an infinite, increasing sequence of r tending to $r=1$. From some r onward $M_2^2(r,f)$ and $M_2^2(r,g)$ are of the form (3.45). Applying Lemma 1 to the difference

$$M_2^2(r,f)-M_2^2(r,g),$$

we find that

$$F(u)\equiv\ldots,\qquad(3.47)$$

where the three dots* stand for the corresponding expression connected with $M_2^2(r,g)$. The general case follows by the substitution $z'=kz$ or $z'=k/z$.

3.5. Lemma 1 gives us still further information. We find, in fact, that (3.47) implies that

$$\left.\begin{array}{ll}\text{(i)} & \text{set of } a\equiv\ldots,\\ \text{(ii)} & \Xi(u-1)\equiv\ldots,\\ \text{(iii)} & \psi(u-1)\equiv\ldots,\\ \text{(iv)} & \cot\pi a\,H_a(u-1)-\tan\pi a\,E_a(u-1)\equiv\ldots.\end{array}\right\}\qquad(3.51)$$

* We use this notation also in the next section.

If none of the differences $|a_\nu - a_\mu|$, $|a_\nu - a_\mu'|$, $|a_\nu' - a_\mu'|$ is equal to $\frac{1}{2}$, E_a and H_a cannot both simultaneously be different from the constant zero, and (3.51) (iv) reduces to

$$H_a \equiv \ldots \qquad E_a \equiv \ldots.$$

It follows now, from (3.51), (3.46) and the corresponding equation connected with $g(z)$, that the equation

$$M_2(r, f) \equiv M_2(r, g) \quad (c < r < 1),$$

implies the equation

$$M_2(r, f) \equiv M_2(r, g) \quad (1 < r < d);$$

i.e., Theorem 2 is proved in this special case.

In the most general case, it follows from the definition of E_a and H_a that (3.51) (iv) can be written in the form

$$u^\gamma \left\{ \cot \pi a \sum_{a_\nu = a + \frac{1}{2}} \int_{\omega_\nu/n}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\frac{d\zeta}{\zeta} - \tan \pi a \sum_{a_\nu = a} \int_{\omega_\nu/u}^{\omega_\nu} h(u\zeta)\bar{h}(1/\zeta)\frac{d\zeta}{\zeta} \right\} \equiv \ldots. \tag{3.52}$$

In this equation we may take u real and greater than 1; the path of integration for a typical integral is then the path BC in the last diagram of §3.3. Then, by (3.13) and (3.41),

$$u^\gamma \int_{\omega_\mu/u}^{\omega_\mu} h(u\zeta)\bar{h}(1/\zeta)\frac{d\zeta}{\zeta} = -u^{\gamma + \Sigma\alpha_\nu} e^{-2\pi\alpha_\mu i} \int_{\omega_\mu/n}^{\omega_\mu} k(u\zeta)\bar{k}(1/\zeta)\frac{d\zeta}{\zeta},$$

where the determinations of $k(z)$ and $\bar{k}(z)$ are such that they are conjugate complex for real $z > 1$. (3.52) becomes, after multiplication by $-e^{2\pi a i}$ $(= -e^{2\pi a' i})$,

$$u^{\gamma + \Sigma\alpha_\nu} \left\{ \cot \pi a \sum_{a_\nu = a + \frac{1}{2}} \int_{\omega_\nu/u}^{\omega_\nu} k(u\zeta)\bar{k}(1/\zeta)\frac{d\zeta}{\zeta} + \tan \pi a \sum_{a_\nu = a} \int_{\omega_\nu/u}^{\omega_\nu} k(u\zeta)\bar{k}(1/\zeta)\frac{d\zeta}{\zeta} \right\} \equiv \ldots. \tag{3.53}$$

The integrand is a regular function of ζ when $1 < |u\zeta| < d$ and $1 < |1/\zeta| < d$. If $1 < |u| < d^2$ these conditions are compatible.

We now study the variation of both sides of (3.53) as u varies. In particular, we start from a positive value of u, $1 < u < d^2$, and let u describe a complete circle $|u| = \text{const.}$ in the positive sense. When we return to the starting point, the path of integration of a typical integral $\int_{\omega_\mu/u}^{\omega_\mu}$ in (3.53) now consists of the circle $|\zeta| = 1/u$ described once in the positive sense and of the straight line from ω_μ/u to ω_μ. The function

$u^{\gamma+\Sigma\alpha_\nu}$ is multiplied by the factor $e^{2\pi(\gamma+\Sigma\alpha_\nu)i}$. The contribution of each circle is

$$e^{2\pi(\gamma+\Sigma\alpha_\nu)i}\,u^{\gamma+\Sigma\alpha}\oint_{|\zeta|=1/u} k(u\zeta)\,\bar{k}(1/\zeta)\,\frac{d\zeta}{\zeta}.$$

By Cauchy's theorem this is equal to

$$e^{2\pi(\gamma+\Sigma\alpha)i}\,u^{\gamma+\Sigma\alpha}\oint_{|\zeta|=1/r} k(u\zeta)\,\bar{k}(1/\zeta)\,\frac{d\zeta}{\zeta} = 2\pi i\,e^{2\pi(\gamma+\Sigma\alpha)i}\,M_2^{\,2}(r,f) \quad (u=r^2).$$

The contribution of the straight lines is the same as in (3.53) but for the factor $e^{2\pi(\gamma+\Sigma\alpha)i}$. Hence, if M_1 is the number of terms of the first sum on the left-hand side of (3.53) and M_2 the number of terms of the second sum, we have

$$2\pi i\,e^{2\pi(\gamma+\Sigma\alpha)i}\,M_2^{\,2}(r,f)\{M_1\cot\pi a+M_2\tan\pi a\}$$

$$+e^{2\pi(\gamma+\Sigma\alpha)i}\;\;\gamma+\Sigma\alpha\;\Big\{\cot\pi a\sum_{\alpha_\nu=a+\frac{1}{2}}\int_{\omega_\nu/u}^{\omega_\nu} k(u\zeta)\,\bar{k}(1/\zeta)\,\frac{d\zeta}{\zeta}$$

$$+\tan\pi a\sum_{\alpha_\nu=a}\int_{\omega_\nu/u}^{\omega_\nu} k(u\zeta)\,\bar{k}(1/\zeta)\,\frac{d\zeta}{\zeta}\Big\}\equiv\ldots. \qquad (3.54)$$

Letting u tend to 1, we have

$$2\pi i\,e^{2\pi(\gamma+\Sigma\alpha)i}(M_1\cot\pi a+M_2\tan\pi a)\,M_2^{\,2}(1,f)\equiv\ldots.$$

But we know that

$$M_2^{\,2}(1,f)=\lim_{\epsilon\to 0}M_2^{\,2}(1-\epsilon,f)=\lim_{\epsilon\to 0}M_2^{\,2}(1-\epsilon,g)=M_2^{\,2}(1,g).$$

Hence
$$e^{2\pi(\gamma+\Sigma\alpha)i}(M_1\cot\pi a+M_2\tan\pi a)=\ldots. \qquad (3.55)$$

And, since $0<a<\frac{1}{2}$, we have $M_1\cot\pi a+M_2\tan\pi a>0$, so that

$$e^{2\pi(\gamma+\Sigma\alpha)i}=\ldots, \qquad (3.56)$$

$$M_2^{\,2}(r,f)\equiv M_2^{\,2}(r,g) \quad (1<r<d)$$

follows now from (3.56), (3.54), (3.53) and (3.55).

4. As $\lambda\to 0$,

$$M_\lambda(r,f)\to\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log|f(re^{i\theta})|\,d\theta\right\}=M_0(r,f)$$

and, as $\lambda\to\infty$,

$$M_\lambda(r,f)\to\max_{|z|=r}|f(z)|=M(r,f).$$

For these two limiting cases Theorem 1 ceases to be true.

For, suppose $f(z)$ to be regular in $|z|<2$ and $f(0)\neq 0$. Then

$$\log M_0(r;f)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log|f(re^{i\theta})|\,d\theta=\log|f(0)|+\sum_{a_n\leqslant r}\log\frac{r}{|a_n|},$$

where the a_n are the zeros of $f(z)$. Obviously we can find two functions regular in $|z|<2$, with the same value at the origin and the same distribution of zeros in $|z|\leqslant 1$, but with zeros of different moduli outside the unit circle. For two such functions we should have

$$M_0(r;f)=M_0(r,g)\quad(r\leqslant 1),$$

$$M_0(r,f)\neq M_0(r,g)\quad(r>r_0>1).$$

To find a Gegenbeispiel in the case $\lambda=\infty$ let

$$f(z)=(z-e_1)(z-e_2),\quad e_2<0<e_1;\ e_1+e_2<0,\quad r_1=\frac{\sqrt{|e_2|}-\sqrt{|e_1|}}{\sqrt{|e_2|}+\sqrt{|e_1|}}\sqrt{|e_1e_2|},$$

$$g(z)=\frac{|e_1|+|e_2|}{2\sqrt{|e_1e_2|}}(z^2+|e_1e_2|),\quad r_2=\frac{1}{r_1}|e_1e_2|.$$

Then*

$$M(r,f)=\begin{cases}|e_1e_2|+(|e_2|-|e_1|)r-r^2 & (0\leqslant r\leqslant r_1),\\ \dfrac{|e_1|+|e_2|}{2\sqrt{|e_1e_2|}}(r^2+|e_1e_2|) & (r_1\leqslant r\leqslant r_2),\\ -|e_1e_2|+(|e_2|-|e_1|)r+r^2 & (r_2\leqslant r).\end{cases}$$

and obviously

$$M(r,g)=\frac{|e_1|+|e_2|}{2\sqrt{|e_1e_2|}}(r^2+|e_1e_2|).$$

Hence $M(r,f)=M(r,g)$ for $r_1\leqslant r\leqslant r_2$, but not for other values of r.

I wish to record my gratitude to Mr. Ingham for his kind advice.

King's College,
Aberdeen.

* See Blumenthal, *Bull. de la Soc. Math. de France*, 35 (1907), 224.

ON THE COEFFICIENTS OF SUBORDINATE FUNCTIONS

By WERNER ROGOSINSKI.

[Received and read 15 December, 1939.]

Introduction.

0.1. Let $\mathfrak{D}$ be a *simply connected* domain on the z-sphere and let $W = F(z)$ be *meromorphic* in $\mathfrak{D}$. We denote by $\mathfrak{G}(F)$ the Riemann domain, extended over the W-sphere, which corresponds to the transform of $\mathfrak{D}$ by $F(z)$†.

Let $w = f(z)$ be a second function, meromorphic in $\mathfrak{D}$. Then $f(z)$ is called *subordinate to* $F(z)$ *in* $\mathfrak{D}$‡, with centre z_0, if

$$f(z_0) = F(z_0) \quad (z_0 \text{ in } \mathfrak{D}), \tag{0.1.1}$$

and if the values w of $f(z)$ in $\mathfrak{D}$, determined by analytical continuation from z_0, are situated in $\mathfrak{G}(F)$. In other words, the Riemann domain $\mathfrak{G}(f)$ has a common point W_0 with $\mathfrak{G}(F)$ and is extended (but is not necessarily schlicht) over $\mathfrak{G}(F)$.

We express this symbolically by

$$f(z) \prec F(z) \text{ in } \mathfrak{D}, \tag{0.1.2}$$

and $F(z)$ is then called *superordinate* to $f(z)$.

It follows from this definition that, if W^* is an inner branch point of order k of $\mathfrak{G}(F)$, and if W^* is also a value of $f(z)$, then it is an inner branch point of order kl, $l \geqslant 1$, of $\mathfrak{G}(f)$. Thus, for instance, $z^{2k} \prec z^2$ in $|z| < 1$ (with centre 0), but z^{2k+1} is not subordinate to z^2.

$f(z)$ is called *schlicht subordinate*§ to $F(z)$ in $\mathfrak{D}$, with centre z_0, if (0.1.1) is satisfied and if $\mathfrak{G}(f)$ is a schlicht part of $\mathfrak{G}(F)$. We express this

† In $\mathfrak{G}(F)$ the inverse function $F^{-1}(W)$ is one-valued. If $F(z)$ takes the value W^* somewhere in $\mathfrak{D}$ with multiplicity $k > 1$, then W^* is an inner branch point of order k of $\mathfrak{G}(F)$. There may be further branch points on the boundary.

‡ Littlewood. Here subordination is restricted to regular and locally schlicht functions $F(z)$. It is the assumption locally schlicht, not that of regularity, that is of special value for the applications.

§ Rogosinski, **5**.

symbolically by

$$f(z) \underset{s}{\prec} F(z) \quad \text{in} \quad \mathfrak{D}. \tag{0.1.3}$$

0.2. The theory of subordination, which is a straightforward generalisation of the principle of maximum modulus, is of great importance in the general theory of analytic functions. Its systematic handling is usually called the *Lindelöf principle*†. There are, however, two quite distinct directions in which this principle can be applied.

I. In the first place, we may start with a given function $F(z)$ and discuss the properties of *the class of all functions* $f(z) \prec F(z)$. A typical result in this direction is Schottky's famous theorem‡ on meromorphic functions which fail to assume three given values. The superordinate function, in this example, is an elliptic modular function.

II. Secondly, we may consider an *arbitrary* function $f(z)$ meromorphic in $\mathfrak{D}$. Suppose that from some given property of $f(z)$ we infer a contradiction to an assumed subordination to some suitable function $F(z)$. If then, for instance, $F(z)$ is locally schlicht [*i.e.* if $\mathfrak{G}(F)$ has no inner branch points] we can conclude that $f(z)$ assumes values outside $\mathfrak{G}(F)$. A typical example in this direction is Landau's theorem in the theory of Picard's "great theorem"†, the function $F(z)$ being again a modular function.

There is no essential difference, of course, between the results in either direction; none, for instance, between Schottky's and Landau's theorems. Generally, however, it can be said that the point of view (II) of the Lindelöf principle is the more important one.

0.3. If $\mathfrak{D}$ has only one boundary point we shall see that subordination is possible only if $f(z)$ and $F(z)$ are essentially the same functions. If $\mathfrak{D}$ has more than one boundary point there is no loss of generality in assuming that $\mathfrak{D}$ is the unit circle $|z| < 1$ and $z_0 = 0$. We may further suppose that $F(0) = 0$.

This, indeed, is only a normalisation of our conditions, and in all that follows we make these assumptions. The notation

$$f(z) \prec F(z) \tag{0.3.1}$$

† Lindelöf. Compare Koebe, Littlewood, Rogosinski, **3, 4.**

‡ See Landau, 103 and 105.

is to be understood to mean that $f(z) \prec F(z)$ in $|z| < 1$, with centre 0, and that

$$f(0) = F(0). \tag{0.3.2}$$

The schlicht subordination $f(z) \prec\prec F(z)$ has a similar meaning.

The "ordinary" subordination (0.3.1) is equivalent to the relation

$$f(z) = F\big(\phi(z)\big), \tag{0.3.3}$$

where $\phi(z)$ is regular in $|z| < 1$ and there satisfies

$$\phi(0) = 0, \quad |\phi(z)| < 1. \tag{0.3.4}$$

If $f(z) \prec\prec F(z)$, then $\phi(z)$ is also schlicht. In fact, by definition, the function $\phi(z) = F^{-1}\big(f(z)\big)$ is one-valued and satisfies (0.3.4), and it is schlicht in the case of schlicht subordination.

0.4. We invariably write

$$F(z) = A_1 z + A_2 z^2 + \ldots + A_n z^n + \ldots, \tag{0.4.1}$$

$$f(z) = a_1 z + a_2 z^2 + \ldots + a_n z^n + \ldots, \tag{0.4.2}$$

these expansions being convergent near $z = 0$. We are concerned in this paper with the relations between the coefficients a_n and A_n involved in the subordination $f(z) \prec F(z)$. If we take the point of view (II) of the Lindelöf principle we arrive at typical "Weierstrass problems", namely the inference from known properties of the a_n, or even of some a_n, of conclusions concerning the values of $f(z)$.

The simplest relation of this kind is the well-known inequality

$$|a_1| \leqslant |A_1|, \tag{0.4.3}$$

which follows immediately from (0.3.3) since

$$(a) \qquad |f'(0)| = |F'(0)\phi'(0)| \leqslant |F'(0)|.$$

We see, moreover, that $|a_1| = |A_1|$ is possible only if $|\phi'(0)| = 1$; and then $\phi(z) = \epsilon z$, $|\epsilon| = 1$, and $f(z) = F(\epsilon z)$.

In spite of the rather obvious character of (0.4.3) the applications of this inequality are very important†. Both Schottky's and Landau's theorems are immediate consequences of (0.4.3).

Let us assume that $A_1 \neq 0$. It is then easy to give conditions on a_n and A_n necessary and sufficient for the subordination $f(z) \prec F(z)$. In

† Compare, for instance, Rogosinski, 3.

fact, the relation $\phi(z) = F^{-1}\big(f(z)\big)$, where $F^{-1}(W)$ is regular near $W = 0$, enables us to express the coefficients of the power series representing $\phi(z)$ by those of $F^{-1}(W)$ and $f(z)$, *i.e.* by the a_n and A_n. On the other hand, necessary and sufficient conditions for coefficients of functions $\phi(z)$ are well known†, whence we get the desired relations. These relations consist of a sequence of inequalities, the n-th inequality involving the coefficients a_k, A_k with $1 \leqslant k \leqslant n$‡.

The first of these inequalities is actually (0.4.3); the second is§

$$(0.4.4) \qquad |A_1^2 a_2 - A_2 a_1^2| \leqslant |A_1|\,(|A_1|^2 - |a_1|^2).$$

The general form of these inequalities, however, is so complicated that they are impracticable for most applications.

0.5. What is wanted for applications are estimates of $|a_n|$ by means of the A_k. Further, best possible results in this direction would provide us with best possible results in the problems of the Lindelöf principle.

We have to distinguish between two main problems.

A. *Given $F(z)$, we want an estimate for $|a_n|$ for fixed n and the whole class of functions $f(z) \prec F(z)$.*

B. *Given $F(z)$ and a fixed function $f(z) \prec F(z)$, we want estimates for the sequence $\{|a_n|\}$.*

In each problem there are two cases corresponding to ordinary and schlicht subordination.

The following two examples illustrate the situation:

I. $F(z) = z$. Here $f(z) \prec z$ means simply $f(0) = 0$ and $|f(z)| < 1$.

The sequence $\{A_n\}$ is 1, 0, 0, ... and we have, in problem A, $|a_n| \leqslant 1 = A_1$ for ordinary subordination, and $|a_n| \leqslant 1/\sqrt{n}$ for schlicht subordination.

In problem B we have, in the case of ordinary subordination, $\Sigma |a_n|^2 \leqslant 1$, and so $a_n = o(1)$ as n tends to infinity. In the case of schlicht subordination $f(z)$ is necessarily schlicht, so that $\Sigma n|a_n|^2 \leqslant 1$ and $a_n = o(n^{-\frac{1}{2}})$ as n tends to infinity.

---

† I. Schur.

‡ Rogosinski, **3**.

§ Rogosinski, **3**, 5; if $f(z) \underset{s}{\prec} F(z)$ we have

$$|A_1^2 a_2 - A_2 a_1^2| \leqslant 2|A_1 a_1|(|A_1| - |a_1|).$$

Rogosinski, **5**, 13.

II. $F(z) = \dfrac{z}{1-z}$. Here $f(z) \prec \dfrac{z}{1-z}$ means $f(0) = 0$ and $\Re f(z) > -\frac{1}{2}$.

The sequence $\{A_n\}$ is 1, 1, 1, ..., and we have in each problem $|a_n| \leqslant 1 = A_n$ as best possible estimate†.

Very little is known about this problem of coefficients‡. But one obvious estimate for problem B can be given at once, namely

$$\overline{\lim}\, \sqrt[n]{|a_n|} \leqslant \max\,(1, \overline{\lim}\, \sqrt[n]{|A_n|}). \tag{0.5.1}$$

In fact, either $F(z)$ is regular for $|z| < 1$, in which case the same holds for $f(z)$, and both the upper limits in (0.5.1) are at most 1; or the radius of convergence P of (0.4.1) is less than 1. Then it follows from (0.3.3) and Schwarz's lemma that the radius of convergence ρ of (0.4.2) is not less than P.

0.6. In the first section of this paper we deal with the problem of coefficients quite generally. The method is a direct one. If we write

$$\phi^k(z) = a_k^{(k)} z^k + a_{k+1}^{(k)} z^{k+1} + \ldots + a_n^{(k)} z^n + \ldots \quad (|z| < 1), \tag{0.6.1}$$

then it follows from (0.3.3) that

$$a_n = \sum_{k=1}^{n} A_k a_n^{(k)}. \tag{0.6.2}$$

The fundamental lemma for dealing with this expression is the inequality

$$|P_n(z)| = \left| \sum_{k=1}^{n} a_n^{(k)} z^k \right| \leqslant 1 \quad (|z| \leqslant 1). \tag{0.6.3}$$

From this we infer that

$$|a_n| \leqslant \sqrt{n} \max_{1 \leqslant k \leqslant n} (|A_k|); \tag{0.6.4}$$

and it can be proved that the order $\sqrt{n}$ of this estimate is in general *best possible* in both problems A and B of § 0.5, and even in the case of schlicht subordination. Since the proof of this is rather elaborate, it is postponed to § 3.

Of the other results in this general problem we mention here the inequalities§

$$\sum_{1}^{m} |a_n|^2 \leqslant \sum_{1}^{m} |A_n|^2; \quad m = 1, 2, \ldots, \tag{0.6.5}$$

† Carathéodory; see also Pólya-Szegö, III, 235.

‡ There is Littlewood's method of mean values mentioned in § 0.7.

§ Rogosinski, **5**, 2. (0.6.5) is a generalization of Littlewood's inequality (0.7.3).

and, in the case of schlicht subordination,

$$\sum_{n=1}^{m} n|a_n|^2 \leqslant \sum_{n=1}^{m} n|A_n|^2 \quad (m = 1, 2, \ldots). \tag{0.6.6}$$

The most important result with regard to applications is given by Theorem VII below:

If the numbers $A_1, A_2, \ldots, A_n$ *are non-negative, non-increasing, and convex, then* $|a_n| \leqslant A_1$.

If the numbers $A_1, A_2, \ldots, A_n$ *are non-negative, non-decreasing, and convex, then* $|a_n| \leqslant A_n$.

Both estimates are best possible. If, for instance, $f(z) \prec \frac{z}{(1-z)^2}$, then $A_n = n$, and we have† $|a_n| \leqslant n$.

0.7. In §2 we discuss subordination to schlicht functions $F(z)$. We here use in part Littlewood's method. Littlewood's main contribution to the theory of subordination is the inequality‡

$$M_q(f, r) \leqslant M_q(F, r) \quad (0 \leqslant q \leqslant \infty, \ 0 \leqslant r < 1), \tag{0.7.1}$$

for the mean values

$$\begin{cases} M_q(f, r) = \left(\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} |f(re^{i\vartheta})|^q \, d\vartheta\right)^{1/q} \quad (0 < q < \infty), \\ M_0(f, r) = \lim\limits_{q \to +0} M_q(f, r) = \exp\left(\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} \log|f(re^{i\vartheta})| \, d\vartheta\right), \\ M_\infty(f, r) = \lim\limits_{q \to \infty} M_q(f, r) = \max\limits_{0 \leqslant \vartheta \leqslant 2\pi} |f(re^{i\vartheta})|. \end{cases} \tag{0.7.2}$$

In (0.7.1) it is assumed that $F(z)$, and thus $f(z)$, is regular in $|z| < 1$. If $q = 2$, we have

$$\sum_{n=1}^{\infty} |a_n|^2 r^{2n} \leqslant \sum_{n=1}^{\infty} |A_n|^2 r^{2n}, \tag{0.7.3}$$

while $q = 1$ provides the estimate

$$|a_n| \leqslant r^{-n} M_1(f, r) \leqslant r^{-n} M_1(F, r). \tag{0.7.4}$$

We shall see that it is possible to prove (0.7.4) directly from (0.6.3) without using the deeper result (0.7.1) for $q = 1$. Taking $r = 1 - 1/n$,

† This example is known. Compare Littlewood, 493.

‡ Littlewood, Theorem 2.

we get† from (0.7.4)

$$|a_n| < e\left[\frac{M_1(F, r)}{r}\right]_{r=1-1/n}. \tag{0.7.5}$$

For certain classes of superordinate functions $F(z)$ this estimate, or a suitable modification, is very useful. On the other hand, this method of mean values cannot yield best possible estimates for the coefficients and, moreover, it does not apply to the general case of a meromorphic function $F(z)$.

If $F(z)$ is schlicht in $|z| < 1$, then Littlewood‡ has proved that

$$M_1(F, r) \leqslant \frac{|A_1| r}{1-r}. \tag{0.7.6}$$

Hence (0.7.5) gives in this case§

$$|a_n| < e|A_1| n. \tag{0.7.7}$$

If $F(z)$ is schlicht, then the so-called "Bieberbach conjecture"

$$|A_n| \leqslant n|A_1|, \tag{0.7.8}$$

with equality (for $n > 1$) only in the case

$$F(z) = A_1 \epsilon^{-1} K(\epsilon z); \quad K(z) = \frac{z}{(1-z)^2} \quad (|\epsilon| = 1), \tag{0.7.9}$$

is still unproved except for $n = 1, 2, 3$‖. But this conjecture has been proved in two special cases: (i) if $F(z)$ has real coefficients¶, and (ii) if $F(z)$ transforms the unit circle in a domain $\mathfrak{G}(F)$ "starlike with respect to 0"††.

A more general proposition, which we may call the "generalised Bieberbach conjecture", is that $f(z) \prec F(z)$, where $F(z)$ is schlicht, implies that

$$|a_n| \leqslant n|A_1|. \tag{0.7.10}$$

This is known to be true for $n = 1, 2$‡‡.

---

† Littlewood, Theorem 3.
‡ Littlewood, Theorem 17.
§ Littlewood, Theorem 1.
‖ Bieberbach, 85. Löwner.
¶ Rogosinski, 2. Dieudonné.
†† Bieberbach, 94. Pólya-Szegö IV, 161.
‡‡ Littlewood, Theorem 16.

The main result of this section is that (0.7.10) *is true for the same two special classes* (i) *and* (ii) *above of superordinate functions* $F(z)$. The sign of equality in (0.7.10) can also be discussed in these two cases.

1. *The general problem.*

1.1. Let $f(z) \prec F(z)$. Then by (0.3.3) we have, near $z = 0$,

$$(1.1.1) \quad f(z) = F\big(\phi(z)\big) = A_1\phi(z) + A_2\phi^2(z) + \ldots + A_n\phi^n(z) + \ldots.$$

Hence, if we put

$$(1.1.2) \qquad \phi^k(z) = a_k^{(k)} z^k + a_{k+1}^{(k)} z^{k+1} + \ldots + a_n^{(k)} z^n + \ldots,$$

it follows that

$$(1.1.3) \qquad a_n = \sum_{k=1}^{n} a_n^{(k)} A_k.$$

Therefore in problem A of §0.5 only the first n coefficients A_k of $F(z)$ are involved, and it is sometimes convenient to assume that $F(z)$ is the polynomial

$$(1.1.4) \qquad S_n(z) = A_1 z + A_2 z^2 + \ldots + A_n z^n.$$

1.2. In §0.3 we excluded as trivial the case where the domain $\mathfrak{D}$ has only one boundary point. This is justified by the following result:

THEOREM I. *If* $f(z) \prec F(z)$ *in the whole plane, i.e. if* $f(Rz) \prec F(Rz$ *for every* $R > 0$, *then*

$$(1.2.1) \qquad f(z) = F(az) \quad (|a| \leqslant 1).$$

Proof. Using (1.1.3) we have

$$(a) \qquad a_n R^n = \sum_{k=1}^{n} a_n^{(k)} A_k R^k; \quad a_n = A_n a_n^{(n)} + \sum_{k=1}^{n-1} a_n^{(k)} A_k R^{k-n}.$$

Here $a_n^{(k)} = a_n^{(k)}(R)$ depends on R. Since $|a_n^{(k)}| \leqslant 1$, the $\sum_1^{n-1}$ tends to zero as $R \to \infty$. Hence

$$(b) \qquad a_n = \lim_{R\to\infty} A_n a_n^{(n)}(R) = A_n \lim_{R\to\infty} \left(a_n^{(1)}(R)\right)^n = A_n a^n \quad (|a| \leqslant 1).$$

1.3. THEOREM II. (i) *If* $f(z) \prec F(z)$, *then*

$$(1.3.1) \qquad \sum_{n=1}^{m} |a_n|^2 \leqslant \sum_{n=1}^{m} |A_n|^2 \quad (m = 1, 2, \ldots).$$

(ii) *If* $f(z) \underset{s}{\prec} F(z)$, *then*

$$\sum_{n=1}^{m} n|a_n|^2 \leqslant \sum_{n=1}^{m} n|A_n|^2 \quad (m = 1, 2, \ldots). \tag{1.3.2}$$

Proof. (i) We have

$$(a) \qquad S_m\big(\phi(z)\big) = A_1\phi(z) + A_2\phi^2(z) + \ldots + A_m\phi^m(z)$$
$$= a_1 z + a_2 z^2 + \ldots + a_m z^m + \sum_{m+1}^{\infty} b_n z^n.$$

Hence, by the limiting case $r = 1$ of Littlewood's inequality (0.7.3),

$$(b) \qquad \sum_{1}^{m} |a_n|^2 \leqslant \sum_{1}^{m} |a_n|^2 + \sum_{m+1}^{\infty} |b_n|^2 \leqslant \sum_{n=1}^{m} |A_n|^2.$$

(ii) First suppose that $F(z)$ is regular for $|z| < 1$. Since, by the definition of schlicht subordination, the function $\phi(z)$ in the relation $f(z) = F\big(\phi(z)\big)$ is schlicht, it follows that the area of $\mathfrak{S}_r(f)$, the transform by $f(z)$ of $|z| \leqslant r$, $0 < r < 1$, is not greater than the area of $\mathfrak{S}_r(F)$. Hence, if $F(z)$ is regular,

$$\sum_{n=1}^{\infty} n|a_n|^2 r^{2n} \leqslant \sum_{n=1}^{\infty} n|A_n|^2 r^{2n}. \tag{1.3.3}$$

(1.3.2) now follows in exactly the same way as (1.3.1) followed from (0.7.3). *Notice that* (1.3.2) *asserts that, for every* m, *the area of the transform of* $|z| < 1$ *by the* m*-th partial sum* $s_m(z)$ *of* $f(z)$ *is not greater than the corresponding area given by the partial sum* $S_m(z)$ *of* $F(z)$†.

1.4. We make repeated use of the following lemma:

LEMMA 1. *Let*

$$g(z) = 1 + c_1 z + c_2 z^2 + \ldots + c_n z^n + \ldots \tag{1.4.1}$$

be regular for $|z| < 1$, *and let* $\Re g(z) > 0$ *for* $|z| < 1$. *Then*

$$\Lambda = \lambda_1 c_1 + \lambda_2 c_2 + \ldots + \lambda_n c_n \tag{1.4.2}$$

† *If, for a given* m, *equality holds in* (1.3.2), *then either* $f(z) = F(\epsilon z_1)$, $|\epsilon| = 1$, *or* $S_m(z) \equiv 0$. Indeed, our proof shows that then each b_n is zero in (*a*). Hence

$$s_m(z) = S_m\big(\Phi(z)\big),$$

and the above remark on the areas gives immediately $\phi(z) = \epsilon z$, if $S_m(z) \not\equiv 0$. The discussion of the sign of equality in (1.3.1) is somewhat more difficult. The result is: *if, for a given* m, *equality holds in* (1.3.1), *then either* $S_m(z) \equiv 0$ *or* $f(z) = F(\epsilon z^p)$, $|\epsilon| = 1$, $1 \leqslant p \leqslant [m/\mu]$ *where* A_μ *is the last coefficient of* $S_m(z)$ *different from zero.*

is situated in the convex cover of the curve

$$\Lambda(\vartheta) = 2(\lambda_1 e^{i\vartheta} + \lambda_2 e^{i2\vartheta} + \ldots + \lambda_n e^{in\vartheta}). \tag{1.4.3}$$

If Λ *is a boundary point of this convex cover then*

$$g(z) = \sum_{j=1}^{J} \mu_j \frac{1+ze^{i\vartheta_j}}{1-ze^{i\vartheta_j}} \quad (\mu_j \geqslant 0); \quad \sum_{j=1}^{J} \mu_j = 1, \tag{1.4.4}$$

where the ϑ_j, $0 \leqslant \vartheta_j < 2\pi$, *are the solutions of*

$$\Lambda(\vartheta) = \Lambda. \tag{1.4.5}$$

This lemma is known†. A special application is Carathéodory's‡ inequality

$$|c_n| \leqslant 2. \tag{1.4.6}$$

Here, if $c_n = 2$, say, $g(z)$ is necessarily of the form (1.4.4) with

$$\vartheta_j = \exp \frac{2j\pi i}{n} \quad (j = 0, 1, 2, \ldots, n-1; \ J = n).$$

1.5. The following lemma is fundamental for the whole investigation.

LEMMA 2. *If* $|z| \leqslant 1$, *then*

$$|P_n(z)| = \left| \sum_{k=1}^{n} a_n^{(k)} z^k \right| \leqslant 1. \tag{1.5.1}$$

If $|\zeta| = 1$ *and* $P_n(\zeta) = \epsilon$, $|\epsilon| = 1$, *then*

$$\frac{\zeta\phi(\epsilon^{-1/n}z)}{1-\zeta\phi(\epsilon^{-1/n}z)} = \sum_{j=1}^{n} \mu_j \frac{\epsilon_j z}{1-\epsilon_j z} \quad \left(\mu_j \geqslant 0; \ \sum_{j=1}^{n} \mu_j = 1\right), \tag{1.5.2}$$

where $\epsilon_j = \exp(2j\pi i/n)$.

Proof. If $|\zeta| = 1$, we have

$$f^*(z) = \frac{\zeta\phi(z)}{1-\zeta\phi(z)} \prec \frac{\zeta z}{1-\zeta z}. \tag{a}$$

Since the n-th coefficient A_n of $\zeta z(1-\zeta z)^{-1}$ is ζ^n, $P_n(\zeta)$ is, by (1.1.3), the n-th coefficient $a_n{}^*$ of $f^*(z)$. But

$$\Re g(z) = \Re\big(1+2f^*(z)\big) > 0. \tag{b}$$

Hence, by (1.4.6), $|P_n(\zeta)| \leqslant 1$. This implies (1.5.1).

† Rogosinski, 2, 105. The lemma is a transform of Carathéodory's well-known characteristic conditions for the coefficients c_n; Carathéodory.

‡ Carathéodory. Pólya-Szegö, III, 235.

If $P_n(\zeta) = \epsilon$, $|\epsilon| = 1$, then, by Lemma 1,

$$(c) \qquad g(\epsilon^{-1/n} z) = \sum_{j=1}^{n} \mu_j \frac{1+\epsilon_j z}{1-\epsilon_j z},$$

since $g(\epsilon^{-1/n} z)$ has its n-th coefficient equal to 2. This is equivalent to (1.5.2).

Remarks: (i) If $|z| \leqslant 1$,

$$(1.5.3) \qquad |P_n^*(z)| = \left| \sum_{k=1}^{n} a_n^{(n+1-k)} z^k \right| \leqslant 1.$$

For $P_n^*(z) = z^{n+1} P_n(z^{-1})$.

(ii) For $|z| \leqslant 1$, and every pair of integers u and m,

$$(1.5.4) \qquad \left| \sum_{k \equiv u \pmod m}^{n} a_n^{(k)} z^k \right| \leqslant 1; \qquad \left| \sum_{k \equiv u \pmod m}^{n} a_n^{(n+1-k)} z^k \right| \leqslant 1.$$

For, if $|\zeta| = 1$,

$$(d) \qquad |\zeta^{-u} P_n(z\zeta)| = \left| \sum_{k=1}^{n} a_n^{(k)} z^k \zeta^{k-u} \right| \leqslant 1.$$

Apply this for all the m-th unit roots ζ and add. This gives the first inequality (1.5.4). The second follows in the same way from (1.5.3).

(iii) We add a typical application in direction (II) of the Lindelöf principle. Let

$$(1.5.5) \qquad f(z) \prec F(z) = \frac{z}{1-z^2} = z+z^3+z^5+\ldots.$$

Then, by (1.1.3) and (1.5.4),

$$(1.5.6) \qquad |a_n| = \left| \sum_{k \equiv 1 \pmod 2}^{n} a_n^{(k)} \right| \leqslant 1.$$

$F(z)$ maps the unit circle on the W-plane cut along the two pieces of the imaginary axis for which $|W| \geqslant \frac{1}{2}$. Working in the same way with $\epsilon F(z)$, $|\epsilon| = 1$, we deduce:

Any function $f(z)$, for which $|a_n| > 1$ for some n, takes some value w on every ray $\arg w = \vartheta$, $|w| > \frac{1}{2}$, in the w-plane.

1.6. Since $|a_n^{(k)}| \leqslant 1$, (1.1.3) yields the trivial estimate

$$(1.6.1) \qquad |a_n| \leqslant \sum_{1}^{n} |A_k| \leqslant n \max_{1 \leqslant k \leqslant n} |A_k|.$$

Let K_n denote the best possible constant, independent of A_k, in the inequality

$$|a_n| \leqslant K_n \max_{1 \leqslant k \leqslant n} |A_k|. \tag{1.6.2}$$

Obviously

$$K_n = \text{u.b.} \sum_{k=1}^{n} |a_n^{(k)}|, \tag{1.6.3}$$

where the upper bound is taken for all possible $\phi(z)$. Taking $\phi(z) = z$ we observe that $K_n \geqslant 1$.

Clearly, $K_1 = 1$; compare (0.4.3). Further

$$K_2 = 1; \quad |a_2| \leqslant \max(|A_1|, |A_2|). \tag{1.6.4}$$

For

(*a*) $\quad |a_2^{(1)}| + |a_2^{(2)}| = |a_2^{(1)}| + |a_1^{(1)}|^2 \leqslant (1 - |a_1^{(1)}|^2) + |a_1^{(1)}|^2 = 1$†.

It is natural that we should look for an estimate of the type (1.6.2). No general estimate of the type $|a_n| \leqslant C_n |A_n|$ can exist since $F(z^n) \prec F(z)$; and here $a_n = A_1$.

THEOREM III. *If* $f(z) \prec F(z)$, *then*

$$|a_n| \leqslant \sqrt{n} \cdot \max_{1 \leqslant k \leqslant n} |A_k|; \quad K_n \leqslant \sqrt{n}. \tag{1.6.5}$$

Proof. It follows from (1.5.1) that

(*b*) $$\sum_{k=1}^{n} |a_n^{(k)}|^2 \leqslant 1.$$

Hence, by Cauchy's inequality,

(*c*) $$\sum_{k=1}^{n} |a_n^{(k)}| \leqslant \sqrt{n} \cdot \sum_{k=1}^{n} |a_n^{(k)}|^2 \leqslant \sqrt{n}.$$

Thus $K_n \leqslant \sqrt{n}$.

Remarks: (i) Another proof follows from Theorem II. (1.3.1) gives

(*d*) $$|a_n|^2 \leqslant \sum_{k=1}^{n} |A_k|^2,$$

and this implies (1.6.5). This proof uses Littlewood's inequality (0.7.3) and is, therefore, less elementary than the first one.

† Since $\phi(z) \prec z$, the inequality (0.4.4) gives $|a_2^{(1)}| < 1 - |a_1^{(1)}|^2$.

(ii) Similarly, (1.3.2) gives the result:

If $f(z) \underset{s}{\prec} F(z)$, *then*

$$|a_n| \leqslant \sqrt{\left(\frac{n+1}{2}\right)} \max_{1 \leqslant k \leqslant n} |A_k|. \tag{1.6.6}$$

1.7. The order $\sqrt{n}$ in (1.6.5) and (1.6.6) is best possible. If we denote by K_n^* the best possible constant, independent of A_k, in the equality

$$|a_n| \leqslant K_n^* \max_{1 \leqslant k \leqslant n} |A_k| \tag{1.7.1}$$

for schlicht subordination, then

$$K_n^* = \text{u.b.} \sum_{k=1}^{n} |a_n^{(k)}|, \tag{1.7.2}$$

where the upper bound is taken for all schlicht functions $\phi(z)$. Obviously $K_n^* \leqslant K_n$.

THEOREM IV. *K_n^* is not $o(\sqrt{n})$.*

And, what is perhaps more surprising,

THEOREM V. *There exists a function $F(z)$ with coefficients $A_n = O(1)$, and a function $f(z) \underset{s}{\prec} F(z)$, such that a_n is not $o(\sqrt{n})$.*

The lengthy proofs† of these theorems are given in §3.

1.8. We have

$$(a) \quad s_m(z) = \sum_{n=1}^{m} a_n z^n = \sum_{n=1}^{m} z^n \sum_{k=1}^{n} a_n^{(k)} A_k = \sum_{k=1}^{n} A_k \sum_{n=k}^{m} a_n^{(k)} z^n = \sum_{k=1}^{n} A_k s_n^{(k)}(z).$$

Now $|\phi(z)/z|^k \leqslant 1$. Hence‡

$$(b) \qquad |s_m^{(k)}(z)| \leqslant r^k \qquad (|z| = r \leqslant \tfrac{1}{2}),$$

and thus

$$|s_m(z)| \leqslant \sum_{k=1}^{m} |A_k| r^k \quad (|z| = r \leqslant \tfrac{1}{2}). \tag{1.8.1}$$

Further, by a well-known result of Landau§,

$$(c) \qquad \left| \sum_{n=k}^{m} a_n^{(k)} \right| \leqslant G_{m-k},$$

† The idea of these proofs was suggested to me by Prof. J. E. Littlewood.

‡ Rogosinski, **1**, 271. Fejér, 79. Landau, 25.

§ Landau, 26.

where

$$(1.8.2)\quad G_n = 1+(\tfrac{1}{2})^2+\left(\frac{1.3}{2.4}\right)^2+\ldots+\left(\frac{1.3\ldots(2n-1)}{2.4\ldots 2n}\right)^2 \simeq \frac{1}{\pi}\log n.$$

Hence, in generalisation of this,

$$(1.8.3)\quad \left|\sum_{n=1}^{m} a_n\right| \leqslant \sum_{k=1}^{m} |A_k|\, G_{m-k} \leqslant G_m \sum_{k=1}^{m} |A_k| \simeq \frac{1}{\pi}\log m \sum_{k=1}^{m} |A_k|.$$

Finally, by a result of I. Schur§, we have, for all $|z| = r \leqslant 1$,

$$(d)\qquad \frac{1}{M-k+1}\sum_{m=k}^{M} |s_m^{(k)}(z)| \leqslant r^k.$$

Hence

$$(e)\qquad \frac{1}{M}\sum_{m=1}^{M} |s_m(z)| = \frac{1}{M}\sum_{m=1}^{M}\left|\sum_{k=1}^{m} A_k s_m^{(k)}(z)\right| \leqslant \frac{1}{M}\sum_{k=1}^{M} |A_k| \sum_{m=k}^{M} |s_m^{(k)}(z)|$$

implies that

$$(1.8.4)\qquad \frac{1}{M}\sum_{m=1}^{M} |s_m(z)| \leqslant \sum_{k=1}^{M} |A_k|\left(1-\frac{k-1}{M}\right) r^k \quad (|z| = r \leqslant 1),$$

which is the generalisation of the above result of Schur.

1.9. If we write $S_n(z) = \sum_{k=1}^{n} A_k z^k$, and if $P_n(z)$ denotes the polynomial in (1.5.1), then

$$(a)\qquad a_n = \sum_{k=1}^{n} a_n^{(k)} A_k = \frac{1}{2\pi}\int_0^{2\pi} S_n(e^{i\vartheta})\, P_n(e^{-i\vartheta})\, d\vartheta,$$

and hence

$$(1.9.1)\qquad |a_n| \leqslant \frac{1}{2\pi}\int_0^{2\pi} |S_n(e^{i\vartheta})|\, d\vartheta.$$

In this estimate we can replace $S_n(e^{i\vartheta})$ by any power series uniformly convergent on the unit circle, and having the given first n coefficients $A_1, A_2, \ldots, A_n$.

Let $F(z)$ be regular for $|z| < 1$ and let $f(z) \prec F(z)$. Then $f(rz) \prec F(rz)$ for every $0 < r < 1$. If we now apply (1.9.1) to this subordination and replace $S_n(rz)$ by $F(rz)$ we get

$$(1.9.2)\qquad |a_n| \leqslant \frac{1}{2\pi r^n}\int_0^{2\pi} |F(re^{i\vartheta})|\, d\vartheta \quad (0 < r < 1).$$

† I. Schur, 130.

This is Littlewood's estimate (0.7.4) with a proof independent of the deeper relation (0.7.1) for $q = 1$.

If $S_n(z) = U_n(z) + iV_n(z)$, we have, in the same way,

$$(1.9.3) \quad a_n = \frac{1}{\pi}\int_0^{2\pi} U_n(e^{i\vartheta})\, P_n(e^{-i\vartheta})\, d\vartheta, \quad |a_n| \leqslant \frac{1}{\pi}\int_0^{2\pi} |U_n(e^{i\vartheta})|\, d\vartheta;$$

and† if $F(z) = U(z) + iV(z)$ is regular for $|z| < 1$,

$$(1.9.4) \qquad |a_n| \leqslant \frac{1}{\pi r^n}\int_0^{2\pi} |U(re^{i\vartheta})\, d\vartheta \quad (0 < r < 1).$$

Similar formulae hold, of course, with $V_n(z)$ and $V(z)$.

1.10. We now proceed to prove some "best possible" results.

THEOREM VI. *Let $f(z) \prec F(z)$.*

(i) *If $A_1 > 0$ and if there is a function*

$$(1.10.1) \qquad F_1(z) = \tfrac{1}{2}A_1 + A_2 z + \ldots + A_n z^{n-1} + \sum_{n+1}^{\infty} B_k z^{k-1},$$

regular in $|z| < 1$ and satisfying $\Re F_1(z) > 0$, then

$$(1.10.2) \qquad |a_n| \leqslant A_1.$$

Equality is possible only if either $f(z) = F(\epsilon z^n)$, $|\epsilon| = 1$, or if the A_k are of the form

$$(1.10.3) \qquad A_k = \sum_{j=1}^{n-1} \rho_j e^{i(k-1)\vartheta_j} \quad (\rho_j \geqslant 0;\ k = 1, 2, \ldots, n).$$

(ii) *If $A_n > 0$ and if there is a function*

$$(1.10.4) \qquad F_2(z) = \tfrac{1}{2}A_n + A_{n-1} z + \ldots + A_1 z^{n-1} + \sum_{n+1}^{\infty} C_k z^{k-1},$$

regular in $|z| < 1$ and satisfying $\Re F_2(z) > 0$, then

$$(1.10.5) \qquad |a_n| \leqslant A_n.$$

Equality is possible only if either $f(z) = F(\epsilon z)$, $|\epsilon| = 1$, or if the A_k are of the form

$$(1.10.6) \qquad A_k = \sum_{j=1}^{n-1} \rho_j e^{i(n-k)\vartheta_j} \quad (\rho_j \geqslant 0;\ k = 1, 2, \ldots, n).$$

† Littlewood, in his printed but unpublished Lecture Notes, proves that (0.7.1) holds with u and U instead of f and F, provided that $q \geqslant 1$. This implies (1.9.4).

Proof. (i) By Lemma 1, $a_n = \sum_{k=1}^{n} a_n^{(k)} A_k$ is situated in the convex cover of the curve

$$(a) \qquad A_1(a_n^{(1)} + a_n^{(2)} e^{i\vartheta} + \ldots + a_n^{(n)} e^{i(n-1)\vartheta}) = A_1 e^{-i\vartheta} P_n(e^{i\vartheta}).$$

Hence, by Lemma 2, $|a_n| \leqslant A_1$.

If $a_n = A_1$, say, then either $P_n(z) \equiv z$, $a_n^{(1)} = 1$, $\phi(z) = z^n$, $f(z) = F(z^n)$; or, by (1.4.4),

$$(b) \qquad F_1(z) = \tfrac{1}{2} \sum_{j=1}^{J} \rho_j \frac{1+e^{i\vartheta_j} z}{1-e^{i\vartheta_j} z} \quad (\rho_j \geqslant 0;\ \sum_{j=1}^{J} \rho_j = A_1),$$

where

$$(c) \qquad P_n(e^{i\vartheta_j}) = e^{i\vartheta_j} \quad (0 \leqslant \vartheta_j < 2\pi).$$

Hence $J \leqslant n-1$ and the first n of the A_k are of the form (1.10.3), since

$$(d) \quad G(z) = \left(F_1(z) + \tfrac{1}{2}A_1\right) z = A_1 z + A_2 z^2 + \ldots + A_n z^n + B_{n+1} z^{n+1} + \ldots$$

$$= \sum_{1}^{J} \rho_j \frac{z}{1-e^{i\vartheta_j} z}.$$

The corresponding function $f(z) = F\left(\phi(z)\right)$ can be found by means of (1.5.2).

(ii) By Lemma 1, $a_n = \sum_{k=1}^{n} a_n^{(n+1-k)} A_{n+1-k}$ is situated in the convex cover of the curve

$$(e) \qquad A_n(a_n^{(n)} + a_n^{(n-1)} e^{i\vartheta} + \ldots + a_n^{(1)} e^{i(n-1)\vartheta}) = A_n e^{-i\vartheta} P_n^*(e^{i\vartheta}).$$

Hence, by (1.5.3), $|a_n| \leqslant A_n$.

The discussion of the sign of equality is similar to that in (i).

Remark. If $A_k = 0$ for $k \not\equiv u \pmod m$, $1 \leqslant u \leqslant m$, $1 \leqslant k \leqslant n$, then we can replace the conditions (1.10.1) or (1.10.4) for the A_k by corresponding conditions for the A_k, $k \equiv u \pmod m$.

If $n < u$ then $a_n = 0$. If $n \geqslant u$, and $A_{u+\lambda m}$ is the last such A_k with $k \leqslant n$, then (1.10.1) must be replaced by $\Re F_1^*(z) > 0$, where

$$(1.10.7) \quad F_1^*(z) = \tfrac{1}{2}A_u + A_{u+m} z + \ldots + A_{u+\lambda m} z^{\lambda} + \sum_{k=\lambda+1}^{\infty} B_k^* z^k \quad (A_u > 0),$$

and (1.10.2) becomes $|a_n| \leqslant A_u$.

The function (1.10.4) must be replaced in a similar way, and we then get $|a_n| \leqslant A_{u+\lambda m}$. The proof is the same as above, by means of (1.5.4).

1.11. The following particular case of Theorem VI is specially noteworthy.

THEOREM VII. *Let* $f(z) \prec F(z)$.

(i) *If, for* $1 \leqslant k \leqslant n$, *the* A_k *are non-negative, non-increasing, and convex, then*

$$|a_n| \leqslant A_1. \tag{1.11.1}$$

(ii) *If, for* $1 \leqslant k \leqslant n$, *the* A_k *are non-negative, non-decreasing, and convex, then*

$$|a_n| \leqslant A_n. \tag{1.11.2}$$

Proof. It is a known result† (which can easily be proved by two partial summations) that the functions

$$(a) \qquad F_1(z) = \tfrac{1}{2}A_1 + A_2 z + \ldots + A_n z^{n-1} + A_n \sum_{k=n}^{\infty} z^k$$

and

$$(b) \qquad F_2(z) = \tfrac{1}{2}A_n + A_{n-1} z + \ldots + A_1 z^{n-1} + A_1 \sum_{k=n}^{\infty} z^k$$

are either identically zero or have, in the cases (i) and (ii) respectively, positive real parts for $|z| < 1$.

Remarks. (i) *If, for* $1 \leqslant k \leqslant n$, *the* A_k *are convex and* $A_k \geqslant A_\mu$, *then*

$$|a_n| \leqslant A_1 + A_n + |A_\mu| - 2A_\mu. \tag{1.11.3}$$

Proof. Write

$$\begin{aligned}(c) \quad S_n(z) &= A_1 z + A_2 z^2 + \ldots + A_n z^n \\ &= [(A_1 - A_\mu) z + (A_2 - A_\mu) z^2 + \ldots + 0 \cdot z^\mu + \ldots + 0 \cdot z^n] \\ &\quad + [0 \cdot z + \ldots + 0 \cdot z^\mu + (A_{\mu+1} - A_\mu) z^{\mu+1} + \ldots + (A_n - A_\mu) z^n] \\ &\quad + A_\mu [z + z^2 + \ldots + z^n] = S_n^{(1)}(z) + S_n^{(2)}(z) + S_n^{(3)}(z).\end{aligned}$$

By (1.1.3), a_n is linear in the A_k. By (1.11.1) the contribution of $S_n^{(1)}(z)$ to $|a_n|$ is $\leqslant A_1 - A_\mu$; by (1.11.2) that of $S_n^{(2)}(z)$ is $\leqslant A_n - A_\mu$, while that of $S_n^{(3)}(z)$ is $\leqslant |A_\mu|$.

† Fejér.

(ii) If the A_k are concave, consider the relation $-f(z) \prec -F(z)$. The $-A_k$ are convex.

(iii) If the A_n are "finally" convex and non-decreasing, then

$$\overline{\lim_{n\to\infty}} \left|\frac{a_n}{A_n}\right| \leqslant 1. \tag{1.11.4}$$

This follows from (1.11.2) and the fact that

$$\lim_{n\to\infty} |a_n^{(1)}A_1 + a_n^{(2)}A_2 + \ldots + a_n^{(k)}A_k| = 0. \tag{d}$$

(iv) If $A_k = 0$ for $k \not\equiv u \pmod m$, $1 \leqslant k \leqslant n$, corresponding results hold if the remaining A_k, $1 \leqslant k \leqslant n$, are convex; compare the remark on Theorem VI.

1.12. Theorem VII is particularly useful in dealing with problems of the Lindelöf principle. We cannot go into systematic details here, but we add some typical applications.

A. Take $F(z) = K(z) = z/(1-z)^2$. The coefficients $A_k = k$ are convex and increasing.

Therefore, *if* $f(z) \prec z/(1-z)^2$ *then*† $|a_n| \leqslant n$.

By (1.1.3) this is the same as

$$|P_n'(1)| = \left|\sum_{k=1}^{n} a_n^{(k)} k\right| \leqslant n. \tag{1.12.1}$$

This is also a consequence of (1.5.1) since, by Bernstein's inequality‡, if $|z| \leqslant 1$,

$$|P_n'(z)| \leqslant n. \tag{1.12.2}$$

This result, applied to functions subordinate to $\epsilon K(z)$, $|\epsilon| = 1$, gives the result:

Every function $f(z)$ *with* $|a_n| > n$ *for some* n *takes, for any* ϑ, *values* $w = |w|e^{i\vartheta}$ *with* $|w| > \frac{1}{4}$.

B. Take $F(z) = [K(z^m)]^{1/m} = z(1-z^m)^{-2/m}$.

If $m \geqslant 2$, the sequence

$$A_k = \frac{2/m(2/m+1)\ldots(2/m+k-1)}{1.2\ldots k} \tag{a}$$

† Littlewood, 493.

‡ Bernstein and M. Riesz. Equality in (1.12.2) occurs only if $P_n(z) = \epsilon z^n$, $|\epsilon| = 1$.

is non-increasing and convex. In fact,

$$(b) \qquad A_k - A_{k+1} = A_k\left(1-\frac{2/m+k}{k+1}\right) = A_k\,\frac{1-2/m}{k+1} \geqslant 0,$$

$$(c) \quad A_k - 2A_{k+1} + A_{k+2} = A_k\,\frac{1-2/m}{k+1} - A_{k+1}\,\frac{1-2/m}{k+2}$$

$$= A_k(1-2/m)\left[\frac{1}{k+1} - \frac{2/m+k}{(k+1)(k+2)}\right]$$

$$= \frac{A_k}{(k+1)(k+2)}\,(1-2/m)(2-2/m) \geqslant 0.$$

Hence, *if* $f(z) \prec z(1-z^m)^{-2/m}$, $m \geqslant 2$, *then* $|a_n| \leqslant 1$.

Applying this to functions subordinate to $\epsilon F(z)$, $|\epsilon| = 1$, we deduce the result:

Let $f(z)$ *be any function with* $|a_n| > 1$ *for some* n. *If* $m \geqslant 2$, *then, for any real* ϑ, $f(z)$ *assumes values* w, *with* $|w| > 2^{-2/m}$, *on at least one of the* m *rays* $\arg\omega = \vartheta + (2k\pi/m)$, $0 \leqslant k < m$.

C. Let $\gamma \geqslant 0$ and take $F(z) = z(1-\gamma z)^{-1}$. The sequence $A_k = \gamma^{k-1}$ is convex. If $\gamma < 1$, $F(z)$ maps the unit circle on the interior of the circle $\mathfrak{K}_\gamma$ extended over the diameter $-(1+\gamma)^{-1}$, $(1-\gamma)^{-1}$; if $\gamma > 1$, on the exterior of this circle. For $\gamma = 1$, $F(z)$ maps the unit circle on the half-plane $\Re W > -\frac{1}{2}$.

If $\gamma \neq 1$, the centre c and the radius r of $\mathfrak{K}_\gamma$ are given by

$$(d) \qquad c = \frac{\gamma}{1-\gamma^2}, \quad r = \frac{1}{|\gamma^2-1|},$$

and hence

$$(e) \qquad c^2 = \begin{cases} r(r-1) \\ r(r+1) \end{cases}, \quad \text{if} \quad \begin{cases} \gamma < 1 \\ \gamma > 1 \end{cases}; \quad \gamma = \frac{|c|}{r}.$$

Given any circle $\mathfrak{K}$ with centre C and radius $R \neq |C|$, there is a function of the form $\lambda^{-1}F(z)$ which maps the unit circle on the interior of $\mathfrak{K}$ if $R > |C|$, and on the exterior of $\mathfrak{K}$ if $R < |C|$. As for λ we have the relations

$$(f) \qquad c = \lambda C, \quad r = |\lambda|\,R, \quad \gamma = \frac{|C|}{R}.$$

Thus, by (e),

$$(g) \quad |\lambda|\,|C|^2 = \begin{cases} R(|\lambda|\,R-1) \\ R(|\lambda|\,R+1) \end{cases}; \quad |\lambda| = \begin{cases} \dfrac{R}{R^2-|C|^2} \\ \dfrac{R}{|C|^2-R^2} \end{cases} \quad \text{if} \quad \begin{cases} R > |C| \\ R < |C| \end{cases}.$$

We now have the following result:

Let $f(z)$ be any function meromorphic in $|z|<1$, $f(0)=0$. Let $\mathfrak{K}$ be any circle with centre C and radius R.

(i) $R \geqslant |C|$. *If $f(z)$ takes no values outside $\mathfrak{K}$, then*

$$|a_n| \leqslant \frac{R^2-|C|^2}{R}. \tag{1.12.3}$$

(ii) $R \leqslant |C|$. *If $f(z)$ omits the values inside $\mathfrak{K}$, then*

$$|a_n| \leqslant \frac{|C|^2-R^2}{R}\left(\frac{|C|}{R}\right)^{n-1}. \tag{1.12.4}$$

1.13. For fixed n, $1 \leqslant p \leqslant n$, write

$$s_p = \sum_{k=1}^{p} a_n^{(k)}. \tag{a}$$

Then, by Lemma 2 and Landau's inequality†,

$$|s_p| \leqslant G_p \leqslant K \log(p+1); \quad |s_n| \leqslant 1; \quad |s_1| \leqslant 1, \tag{b}$$

where K is an absolute constant. Hence, putting $\Delta_k = A_k - A_{k+1}$, we have

$$a_n = \sum_{k=1}^{n} a_n^{(k)} A_k = \sum_{k=1}^{n-1} s_k \Delta_k + s_n A_n, \tag{c}$$

$$|a_n| \leqslant K \sum_{1}^{n-1} \log(k+1)|\Delta_k| + |A_n|. \tag{1.13.1}$$

Similarly, by means of (1.5.3),

$$|a_n| \leqslant K \sum_{1}^{n-1} \log(k+1)|\Delta_{n-k}| + |A_1|. \tag{1.13.2}$$

THEOREM VIII. *Let $f(z) \prec F(z)$. If $A_1, A_2, \ldots, A_n$ are non-decreasing, then*

$$|a_n| \leqslant K \log n (A_n - A_1) + \min(|A_1|, |A_n|). \tag{1.13.3}$$

The proof follows immediately from (1.13.1) and (1.13.2).

Remarks. (i) If $f(z) \prec F(z)$ and the sequence $\{A_n\}$ is non-decreasing, then

$$|a_n| \leqslant K_1 \log n (A_n - A_1). \tag{1.13.4}$$

† Compare 1.8 (c).

(ii) *If* $f(z) \prec F(z)$, $A_n = O(1)$ *and non-decreasing, then*

$$a_n = o(\log n). \tag{1.13.5}$$

For, since A_n converges, we have in (1.13.1), if $n > m \geqslant M(\epsilon)$,

$$(d) \qquad \sum_{m}^{n-1} \log(k+1)|\Delta_k| \leqslant \log n(A_n - A_m) \leqslant \epsilon \log n.$$

(iii) If $A_k = 0$, $k \not\equiv u \pmod m$, similar results hold if the remaining A_k are non-decreasing.

(iv) One more partial summation applied to (c) would supply a new direct proof of Theorem VII.

2. *Subordination to schlicht functions.*

2.1. Let us first assume only that $F(z)$ is regular in $|z| < 1$. Hence, if $f(z) \prec F(z)$, $f(z)$ is also regular.

We then have Littlewood's inequality (0.7.1) for the mean values $M_q(f, r)$. Moreover, if $f(z) \underset{s}{\prec} F(z)$,

$$\frac{\partial}{\partial r} M_q{}^q(f, r) \leqslant \frac{\partial}{\partial r} M_q{}^q(F, r) \quad (0 < q < \infty,\ 0 < r < 1). \tag{2.1.1}$$

This follows immediately from Prawitz's identity†

$$r \frac{\partial}{\partial r} \int_{-\pi}^{\pi} |f(re^{i\vartheta})|^q d\vartheta = q \int_{|z|=r} R^q d\Phi \quad (0 < q < \infty,\ f(re^{i\vartheta}) = Re^{i\Phi}), \tag{2.1.2}$$

valid for every regular $f(z)$.

In fact, from the definition of schlicht subordination we see that the right-hand side of (2.1.2) for $f(z)$ is not greater than the same expression for $F(z)$.

2.2. Henceforth we assume throughout this section that $F(z)$, but not necessarily $f(z)$, is schlicht in $|z| < 1$. It is then known‡ that

$$|F(z)| \leqslant \frac{|A_1| r}{(1-r)^2} \quad (|z| = r < 1). \tag{2.2.1}$$

---

† Prawitz. (2.1.2) is a simple consequence of the Cauchy-Riemann equation

$$\frac{\partial \Phi}{\partial \vartheta} = \frac{r}{R} \frac{\partial R}{\partial r};$$

(2.1.1) is implicitly stated in Prawitz's paper.

‡ Bieberbach, 89.

Let us assume that

$$|F(z)| \leqslant \frac{Kr}{(1-r)^{\alpha+1}} \quad (-1 < \alpha \leqslant 1), \tag{2.2.2}$$

Formula (2.1.2) then obviously gives

$$\frac{\partial}{\partial r}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |F(re^{i\vartheta})|\, d\vartheta\right) \leqslant \frac{K}{(1-r)^{\alpha+1}}, \tag{a}$$

and thus

$$M_1(F, r) \leqslant \begin{cases} \dfrac{K}{\alpha}\left[\dfrac{1}{(1-r)^{\alpha}} - 1\right] \\ K \log \dfrac{1}{1-r} \end{cases}, \quad \text{if} \quad \begin{cases} \alpha > 0, \\ \alpha = 0. \end{cases} \tag{2.2.3}$$

Hence, from (0.7.5), we get

THEOREM IX. *Let $f(z) \prec F(z)$, where $F(z)$ is schlicht and satisfies* (2.2.2) *with some α, $0 \leqslant \alpha \leqslant 1$. Then*

$$|a_n| < \begin{cases} e\,\dfrac{K}{1-1/n}\,(n^{\alpha}-1) \\ e\,\dfrac{K}{1-1/n}\log n \end{cases}, \quad \text{if} \quad \begin{cases} \alpha > 0, \\ \alpha = 0. \end{cases} \tag{2.2.4}$$

In particular the coefficients A_n themselves satisfy this inequality. Conversely, if

$$A_n = O(n^{\alpha}) \quad (\alpha > -1), \tag{2.2.5}$$

then (2.2.2) is satisfied for some K.

Hence, by Theorem IX, (2.2.5) yields

$$a_n = \begin{cases} O(n^{\alpha}) \\ O(\log n) \end{cases}, \quad \text{if} \quad \begin{cases} \alpha > 0, \\ \alpha = 0. \end{cases} \tag{2.2.6}$$

The interesting question, whether this estimate is best possible in the case $\alpha = 0$, remains unanswered.

It is further known† that (2.2.2) implies

$$A_n = \begin{cases} O(n^{\alpha}) \\ O(n^{-\frac{1}{2}} \log n) \end{cases}, \quad \text{if} \quad \begin{cases} \alpha > -\frac{1}{2}, \\ \alpha = -\frac{1}{2}. \end{cases} \tag{2.2.7}$$

† Spencer, Theorem 3.

The same holds for a_n if $f(z) \underset{s}{\prec} F(z)$, since then $f(z)$ is also schlicht and satisfies (2.2.2).

2.3. THEOREM X. *Let $f(z) \prec F(z)$, where $F(z)$ is schlicht and $\mathfrak{G}(F)$ convex. Then*†

$$|a_n| \leqslant |A_1|. \tag{2.3.1}$$

Equality occurs only either if

$$f(z) = F(\epsilon z^n) \quad (|\epsilon| = 1), \tag{2.3.2}$$

or if $\mathfrak{G}(F)$ is a half-plane and $f(z)$ is of the type (2.3.4) *below.*

Proof. (i) Put $\eta = \exp 2\pi i/n$. Since $\mathfrak{G}(F)$ is convex we have

$$(a) \quad g(z^n) = \frac{f(\eta z)+f(\eta^2 z)+\ldots+f(\eta^n z)}{n} = a_n z^n + a_{2n} z^{2n} + \ldots \prec F(z),$$

$$(b) \quad g(z) = a_n z + a_{2n} z^2 + \ldots \prec F(z).$$

Hence, by (0.4.3), $|a_n| \leqslant |A_1|$.

(ii) The discussion of the sign of equality in (2.3.1) is much more tiresome. Let us assume that $a_n = A_1$, say. Then, again by (0.4.3),

$$(c) \quad g(z) = F(z).$$

Let us further assume that $n > 1$, the case $n = 1$ being covered by (c). By (1.1.3)

$$(d) \quad a_n = \sum_{k=1}^{n} \alpha_n^{(k)} A_k = Q(\alpha_1^{(1)}, \alpha_2^{(1)}, \ldots, \alpha_n^{(1)}),$$

where Q is a certain polynomial in the first n coefficients $\alpha_k^{(1)}$ of $\phi(z)$. The maximum of $|Q|$ for all possible $\phi(z)$ is exactly $|A_1|$, as is shown by (2.3.1) and (2.3.2).

By a theorem of I. Schur‡ this maximum is attained only if $\phi(z)$ is of the form

$$(e) \quad \phi(z) = \epsilon z \prod_{k=1}^{K} \frac{z-a_k}{z\bar{a}_k - 1} \quad (|a_k| < 1, \ |\epsilon| = 1),$$

where $K \leqslant n-1$.

Let $|\zeta| = 1$. The boundary values $F(\zeta)$ exist finitely or infinitely. In the case of a strip $\mathfrak{G}(F)$ the value ∞ occurs twice as a boundary point of $\mathfrak{G}(F)$ and thus as a boundary value $F(\zeta)$.

† (2.3.1) ought to be known, but I have been unable to find it stated explicitly except for special $F(z)$; compare Landau, 34. See also Carathéodory's inequality (1.4.6).

‡ Schur, Satz XIII.

Also $\phi(\zeta)$ exists, and we have $|\phi(\zeta)| = 1$. Thus the boundary values

$$(f) \qquad f(\zeta) = \lim_{z \to \zeta} F\big(\phi(z)\big) = F\big(\phi(\zeta)\big)$$

exist finitely or infinitely, and are situated on the boundary of $\mathfrak{G}(F)$. If ζ moves round the unit circle in the positive sense, $f(\zeta)$ moves $K+1$ times round the boundary of $\mathfrak{G}(F)$ in the same sense. Moreover, by (a) and (c),

$$(g) \qquad \frac{f(\zeta)+f(\eta\zeta)+\ldots+f(\eta^{n-1}\zeta)}{n} = F(\zeta^n).$$

There are three possibilities for given ζ:

α. The n values

$$(h) \qquad f(\zeta), f(\eta\zeta), \ldots, f(\eta^{n-1}\zeta)$$

are equal, finite or infinite, and give the *same* boundary point† $W(\zeta) = F(\zeta^n)$ of $\mathfrak{G}(F)$.

β. One of the values (h) at least is infinite, but they do not‡ all give the same boundary point of $\mathfrak{G}(F)$. *This, by* (f), *can only happen for a finite number of values* ζ.

γ. All the values (h) are finite, and at least two are different. *In this case, by* (g), *the stretch* $S(\zeta)$ *joining two different values* (h) *belongs to the boundary of* $\mathfrak{G}(F)$ *and bears all the values* (h) *and their arithmetic mean* $W(\zeta) = F(\zeta^n)$.

Now let ζ, in this case, move round the unit circle in the positive sense. So long as we are dealing with case (γ) the line $L(\zeta)$, bearing $S(\zeta)$, is obviously unable to change. If ζ passes through a value of behaviour (β), the line $L(\zeta)$ also is unable to change, as is shown by simple geometrical considerations.

Hence the line $L(\zeta)$ *can change only if* ζ *passes through a value of behaviour* (α).

Now (α) implies that $f(\zeta) = f(\eta\zeta)$ and thus $\phi(\zeta) = \phi(\eta\zeta)$. This, by (e), is an equation of degree $2K \leqslant 2(n-1)$ in ζ.

Either (i) this equation is identically satisfied. Then (α) occurs for all ζ and (g) yields $f(z) = F(z^n)$.

Or (ii) there are at most $2(n-1)$ values ζ of behaviour (α). Hence, by the above discussion, $\mathfrak{G}(F)$ is necessarily a "polygonal" domain, bounded by a finite number of lines.

---

† This is noticeable only in the case of a strip $\mathfrak{G}(F)$, $W(\zeta) = \infty$.

‡ In the case of a strip $\mathfrak{G}(F)$ all the values (h) may be infinite without giving the same boundary point of $\mathfrak{G}(F)$.

But, to any boundary point $W(\zeta)$ of type (a), there belong, by definition, exactly n solutions $\zeta, \eta\zeta, \ldots, \eta^{n-1}\zeta$ of the equation. Hence at most one such boundary point $W(\zeta)$ can exist. It follows that $\mathfrak{G}(F)$ is a half-plane, because any other polygonal domain has at least two boundary lines and thus at least two boundary points $W(\zeta)$ of type (a).

It follows that

$$F(z) = \frac{A_1 z}{1-\epsilon z} \quad (|\epsilon| = 1). \tag{2.3.3}$$

If $a_n = A_1 > 0$, say, then $\Re\big(\epsilon f(z)\big) > -\frac{1}{2}A_1$, and thus, by Lemma 1,

$$f(z) = \epsilon^{-1} \sum_{j=1}^{n} \mu_j \frac{\epsilon_j z}{1-\epsilon_j z} \quad (\mu_j \geqq 0,\ \sum_1^n \mu_j = A_1,\ \ \epsilon_j = \epsilon^{1/n} \exp 2j\pi i/n). \tag{2.3.4}$$

(iii) A very short proof of (ii) is practicable if we impose the extra condition

$$\sum_1^\infty |A_k|^2 < \infty. \tag{2.3.5}$$

This is equivalent to the condition that the boundary values $F(\zeta)$ are of class L^2, as is the case for bounded domains $\mathfrak{G}(F)$, and also when $\mathfrak{G}(F)$ is contained in a strip or in an angle of aperture less than $\frac{1}{2}\pi$. We then get from (c)

$$f(z) \prec g(z) = F(z), \tag{i}$$

and thus, by Littlewood's inequality (0.7.3),

$$\sum_1^\infty |a_k|^2 \leqslant \sum_1^\infty |a_{kn}|^2 = \sum_1^\infty |A_k|^2 < \infty. \tag{j}$$

Hence $a_k = 0$ if $k \not\equiv 0 \pmod n$, $f(z) = g(z^n) = F(z^n)$.

2.4. THEOREM XI. *Let $f(z) \prec F(z)$, where $F(z)$ is schlicht and $\mathfrak{G}(F)$ starlike with respect to the origin. Then*

$$|a_n| \leqslant n|A_1|. \tag{2.4.1}$$

If $n > 1$, equality occurs only if

$$F(z) = \frac{A_1 z}{(1-\eta z)^2} \ (|\eta| = 1); \quad f(z) = F(\epsilon z) \ (|\epsilon| = 1). \tag{2.4.2}$$

Proof. (i) If $\mathfrak{G}(F)$ is starlike, and

$$G(z) = \int_0^z \frac{F(u)}{u}\, du = A_1 z + \frac{A_2}{2} z^2 + \ldots + \frac{A_n}{n} z^n + \ldots, \tag{a}$$

then $\mathfrak{G}(G)$ is convex, and conversely†. Consider

$$(b) \qquad g(z) = G\big(\phi(z)\big) = b_1 z + b_2 z^2 + \ldots + b_n z^n + \ldots.$$

Since $g(z) \prec G(z)$, Theorem X gives

$$(c) \qquad |b_n| = \left| A_1 a_n^{(1)} + \frac{A_2}{2} a_n^{(2)} + \ldots + \frac{A_n}{n} a_n^{(n)} \right| \leqslant |A_1|.$$

Replacing $F(z)$ by $F(\epsilon z)$, $|\epsilon| = 1$, we get

$$(d) \qquad |Q_n(z)| = \left| A_1 a_n^{(1)} z + \frac{A_2}{2} a_n^{(2)} z^2 + \ldots + \frac{A_n}{n} a_n^{(n)} z^n \right| \leqslant |A_1|$$

for all $|z| \leqslant 1$. Hence, by Bernstein's inequality (1.12.2),

$$(e) \qquad |Q_n'(1)| = |A_1 a_n^{(1)} + A_2 \xi_n^{(2)} + \ldots + A_n a_n^{(n)}| = |a_n| \leqslant n |A_1|.$$

(ii) Let $a_n = nA_1$, say. Then $Q_n'(1) = nA_1$. This is possible only‡ if $Q_n(z) = A_1 z^n$. Hence $a_n^{(k)} = 0$ if $k < n$, and $a_n^{(n)} A_n = nA_1$. It follows that $|a_n^{(n)}| = |a_1^{(1)}|^n = 1$, since $|A_n| \leqslant n|A_1|$ by (2.4.1). Hence $\phi(z) = \epsilon z$. Further, by (c), $b_n = A_1$ and thus, by Theorem X,

$$(f) \qquad G(z) = \frac{A_1 z}{1 - \eta z}; \quad F(z) = zG'(z) = \frac{A_1 z}{(1-\eta z)^2} \quad (|\eta| = 1).$$

2.5. THEOREM XII. *Let $f(z) \prec F(z)$, where $F(z)$ is schlicht and has real coefficients. Then (2.4.1) holds. If $n > 1$ equality occurs only in the case (2.4.2), $\eta = \pm 1$.*

Proof. (i) We prove, more generally, *that (2.4.1) holds if $F(z)$ is "typically real"§, i.e.*, putting $F(re^{i\vartheta}) = U(re^{i\vartheta}) + iV(re^{i\vartheta})$, $0 < r < 1$, if

$$(a) \qquad V(re^{i\vartheta}) \geqslant 0 \quad (0 \leqslant \vartheta \leqslant \pi); \quad V(re^{i\vartheta}) = -V(re^{-i\vartheta}).$$

† Pólya-Szegö, III, 110. The proof is based on the theorem that necessary and sufficient conditions for $\mathfrak{G}(F)$ to be starlike and $\mathfrak{G}(G)$ to be convex are the inequalities

$$(*) \quad \Re z \frac{F'(z)}{F(z)} > 0; \qquad (**) \quad \Re z \frac{G''(z)}{G'(z)} > -1 \quad (|z| < 1).$$

But neither the statements of Pólya-Szegö nor those of Bieberbach, 92–93, seem to be sufficient to prove the *necessity* of this. In the case of a convex $\mathfrak{G}(G)$ a proof can easily be obtained by approximating to $\mathfrak{G}(G)$ by convex polygons. For the corresponding functions G the Schwarz-Christoffel formula gives (**) immediately. For starlike $\mathfrak{G}(F)$ the proof is less obvious. It follows from the identity

$$(***) \quad \Re z \frac{F'(z)}{F(z)} = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(\phi - \vartheta) + r^2} d\Phi(\vartheta) > 0 \quad (z = re^{i\phi},\ 0 < r < 1),$$

where $\Phi(\vartheta) = \arg F(e^{i\vartheta})$. The boundary values $F(e^{i\vartheta})$ exist for almost all ϑ.

‡ See footnote ‡, p. 65.

§ Rogosinski, 2.

The A_k, of course, are real. The condition (a) is satisfied if $F(z)$ is schlicht and has real coefficients.

Changing $f(z)$ into $f(\epsilon z)$ with a suitable ϵ, $|\epsilon| = 1$, we may assume that a_n is real. We then have

$$(b) \qquad a_n = \sum_{k=1}^{n} a_n^{(k)} A_k = \sum_{k=1}^{n} u_n^{(k)} A_k \quad (u_n^{(k)} = \Re a_n^{(k)}).$$

It is known† that a_n is situated on the interval

$$(c) \qquad \Lambda(\vartheta) = A_1 \sum_{k=1}^{n} u_n^{(k)} \frac{\sin k\vartheta}{\sin \vartheta} \quad (0 \leqslant \vartheta \leqslant \pi).$$

Now Lemma 2, *i.e.*, $|P_n(z)| \leqslant 1$, implies that

$$(d) \qquad \left| \sum_{k=1}^{n} u_n^{(k)} \sin k\vartheta \right| \leqslant 1,$$

and from this it follows, by a theorem of M. Riesz‡, that

$$(e) \qquad \left| \sum_{k=1}^{n} u_n^{(k)} \frac{\sin k\vartheta}{\sin \vartheta} \right| \leqslant n.$$

Hence $|a_n| \leqslant n|A_1|$.

(ii) Let $n > 1$, $a_n = nA_1 > 0$, say. Then, again by M. Riesz's theorem,

$$(f) \qquad \sum_{k=1}^{n} u_n^{(k)} \sin k\vartheta = \pm \sin n\vartheta; \quad \phi^n(z) = \pm z^n.$$

In the case of the positive sign, we have $\Lambda(\vartheta) = A_1 \sin n\vartheta / \sin \vartheta$.

Now $\Lambda(\vartheta) = n$ gives $\vartheta = 0$ and, if n odd, also $\vartheta = \pi$. Hence, by the second part of the result on typically real functions quoted above,

$$(2.5.1) \quad F(z) = \mu_1 \frac{z}{(1-z)^2} + \mu_2 \frac{z}{(1+z)^2}; \quad \mu_1, \mu_2 \geqslant 0; \quad \mu_1 + \mu_2 = A_1.$$

Furthermore $\phi(z) = \epsilon z$, $\epsilon^n = 1$.

In the other case $\Lambda(\vartheta) = -A_1 \sin n\vartheta / \sin \vartheta = n$ gives $\vartheta = \pi$, if n is even. Hence again $F(z)$ is of type (2.5.1) with $\mu_1 = 0$. We now have $\phi(z) = \epsilon z$, $\epsilon^n = -1$.

Of the functions (2.5.1) only $F(z) = A_1 z/(1 \pm z)^2$ are schlicht. This proves the second part of the theorem.

† Rogosinski, **2**, 116.

‡ M. Riesz. See also Pólya-Szegö, VI, 81,

3. *Proofs of Theorems IV and V.*

3.1. In order to construct the examples proving Theorems IV and V, we require some lemmas concerning integrals of the type

$$I_n = \int_a^b \cos\left(n\psi(t)\right) dt. \tag{3.1.1}$$

LEMMA 3. *If, in* $\langle a, b\rangle$, $\psi'(t)$ *exists and is positive and non-increasing, then*

$$|I_n| \leqslant \frac{2}{n\psi'(b)}. \tag{3.1.2}$$

Proof. We have, by the second mean value theorem, for some ξ in $\langle a, b\rangle$

$$(a) \qquad I_n = \int_a^b \cos\left(n\psi(t)\right) \frac{\psi'(t)}{\psi'(t)}\, dt = \frac{1}{\psi'(b)} \int_\xi^b \cos\left(n\psi(t)\right) \psi'(t)\, dt$$

$$= \frac{1}{\psi'(b)} \frac{\sin\left(n\psi(t)\right)}{n}\Bigg|_\xi^b.$$

This gives (3.1.2).

COROLLARY. *If* $\psi(t) = \psi(t, \alpha)$ *satisfies the conditions of Lemma* 3 *for each value of the parameter* α *under consideration, and if* $\psi'(b, \alpha) \geqslant d > 0$ *for all* α, *then*

$$I_n = O(1/n), \tag{3.1.3}$$

uniformly in α.

3.2. LEMMA 4. *If, in* $\langle a, b\rangle$, $\psi'''(t)$ *exists and is continuous, and if* $\psi''(t) > 0$ *and* $\psi'(t)$ *vanishes at exactly one point* τ, $a < \tau < b$, *then*

$$I_n = \sqrt{\left(\frac{2}{n\psi''(\tau)}\right)}\left[\cos\left(n\psi(\tau)\right) \int_0^\infty \frac{\cos u}{\sqrt{u}}\, du \right.$$

$$\left. -\sin\left(n\psi(\tau)\right) \int_0^\infty \frac{\sin u}{\sqrt{u}}\, du\right] + O(n^{-\frac{3}{2}}). \tag{3.2.1}$$

Proof. We have, for small $|t-\tau|$,

$$(a) \qquad \psi(t) = \psi(\tau) + \frac{(t-\tau)^2}{2}\psi''(\tau) + O(|t-\tau|^3),$$

$$(b) \qquad \psi'(t) = (t-\tau)\psi''(\tau) + O\left((t-\tau)^2\right); \quad \frac{1}{\psi'(t)} = O\left(\frac{1}{|t-\tau|}\right).$$

Now, for sufficiently large n, put

$$(c)\quad I_n=\int_a^{\tau-n^{-\frac{2}{5}}}+\int_{\tau-n^{-\frac{2}{5}}}^{\tau+n^{-\frac{2}{5}}}+\int_{\tau+n^{-\frac{2}{5}}}^{b}\cos\left(n\psi(t)\right)dt=A_n+B_n+C_n.$$

By Lemma 3,

$$(d)\qquad |A_n|\leqslant\frac{2}{n|\psi'(\tau-n^{-\frac{2}{5}})|}=O\left(\frac{1}{n\,.\,n^{-\frac{2}{5}}}\right)=O(n^{-\frac{3}{5}}).$$

Similarly $C_n=O(n^{-\frac{3}{5}})$. Finally

$$(e)\quad B_n=\int_{\tau-n^{-\frac{2}{5}}}^{\tau+n^{-\frac{2}{5}}}\cos\left[n\left(\psi(\tau)+\tfrac{1}{2}(t-\tau)^2\psi''(\tau)+O(|t-\tau|^3)\right)\right]dt$$

$$=\int_{\tau-n^{-\frac{2}{5}}}^{\tau+n^{-\frac{2}{5}}}\cos\left[n\left(\psi(\tau)+\tfrac{1}{2}(t-\tau)^2\psi''(\tau)\right)\right]dt+O\left(n\int_{\tau-n^{-\frac{2}{5}}}^{\tau+n^{-\frac{2}{5}}}|t-\tau|^3\,dt\right)$$

$$=\sqrt{\left(\frac{2}{n\psi''(\tau)}\right)}\int_0^{\frac{1}{2}\psi''(\tau)n^{\frac{1}{5}}}\frac{\cos\left(n\psi(\tau)+u\right)}{\sqrt{u}}\,du+O(n^{-\frac{3}{5}}).$$

But

$$(f)\quad \int_0^{\frac{1}{2}\psi''(\tau)n^{\frac{1}{5}}}\frac{\cos\left(n\psi(\tau)+u\right)}{\sqrt{u}}\,du=\int_0^{\infty}\frac{\cos\left(n\psi(\tau)+u\right)}{\sqrt{u}}\,du+O(n^{-\frac{1}{10}})$$

$$=\cos\left(n\psi(\tau)\right)\int_0^{\infty}\frac{\cos u}{\sqrt{u}}\,du-\sin\left(n\psi(\tau)\right)\int_0^{\infty}\frac{\sin u}{\sqrt{u}}\,du+O(n^{-\frac{1}{10}}).$$

This, together with (e), completes the proof.

COROLLARY. *If $\psi(t)=\psi(t,\alpha)$ satisfies the conditions of Lemma 4 for every value of the parameter α that occurs, and if $\psi''(\tau)=\psi''(\tau_\alpha)\geqslant d>0$, $|\psi'''(\tau_\alpha)|\leqslant C$, uniformly for all α, then (3.2.1), with $\tau=\tau_\alpha$, holds uniformly in α.*

This follows immediately from the above proof.

3.3. Let $\rho>1$ and

$$(3.3.1)\qquad \Phi(z)=\Phi_\rho(z)=K^{-1}\left(\frac{K(z)}{\rho}\right);\quad K(z)=\frac{z}{(1-z)^2}.$$

$\Phi(z)$ maps the unit circle on the unit circle slit along a certain interval $\langle-\pi,\,-d(\rho)\rangle$. Furthermore, $\Phi(0)=0$†.

† For dealing with Theorem IV, Prof. J. E. Littlewood suggested that I should try a function similar to that in (3.3.1).

If $z = e^{i\vartheta}$, $|\vartheta| \leqslant \pi$, then $\Phi(e^{i\vartheta}) = e^{i\chi(\vartheta)}$ if $|\vartheta| \leqslant \vartheta_0$, where

$$(a) \qquad K(e^{i\vartheta_0}) = -\tfrac{1}{4}\rho; \quad \frac{e^{i\vartheta_0}}{(1-e^{i\vartheta_0})^2} = -\tfrac{1}{4}\rho; \quad \sin\tfrac{1}{2}\vartheta_0 = 1/\sqrt{\rho}.$$

If $\vartheta_0 < |\vartheta| \leqslant \pi$, then $\Phi(e^{i\vartheta}) = -\eta(\vartheta)$, $d(\rho) \leqslant \eta < 1$, where

$$(b) \qquad K(-d) = \frac{K(-1)}{\rho}; \quad \frac{d}{(1+d)^2} = \frac{1}{4\rho}.$$

Put $\Phi^k(z) = \sum_{n=k}^{\infty} \alpha_n^{(k)} z^n$. Then $\alpha_n^{(k)}$ is real, and we have

$$(3.3.2) \qquad \alpha_n^{(k)} = \frac{2}{\pi}\left\{\int_0^{\vartheta_0} \cos\left(k\chi(\vartheta)\right)\cos n\vartheta\, d\vartheta + (-1)^k \int_{\vartheta_0}^{\pi} \eta^k(\vartheta)\cos n\vartheta\, d\vartheta\right\}.$$

We next discuss the behaviour of $\alpha_n^{(k)}$, where

$$(3.3.3) \qquad k = an, \quad a > 0.$$

3.4. We first prove, for fixed ρ, that

$$(3.4.1) \qquad \int_{\vartheta_0}^{\pi} \eta^{an}(\vartheta)\cos n\vartheta\, d\vartheta = O\left(\frac{1}{n}\right),$$

uniformly for all $a \geqslant a_0 > 0$.

Let $\vartheta_0 < \vartheta < \pi$. Then

$$(a) \qquad K(-\eta) = \frac{1}{\rho}K(e^{i\theta}); \quad \frac{-\eta}{(1+\eta)^2} = \frac{1}{\rho}\frac{e^{i\vartheta}}{(1-e^{i\vartheta})^2} = \frac{1}{\rho(2i\sin\frac{1}{2}\vartheta)^2},$$

$$(b) \qquad \eta^{\frac{1}{2}} + \eta^{-\frac{1}{2}} = 2\sqrt{\rho}\sin\tfrac{1}{2}\vartheta,$$

$$(c) \qquad \tfrac{1}{2}\eta'(\eta^{-\frac{1}{2}} - \eta^{-\frac{3}{2}}) = \tfrac{1}{2}\eta'\eta^{-\frac{1}{2}}(1-\eta^{-1}) = \sqrt{\rho}\cos\tfrac{1}{2}\vartheta.$$

Hence $\eta' < 0$ and $\eta'(\pi) = 0$, $\eta'(\vartheta_0) = -\infty$.

Choose ϑ_1, $\vartheta_0 < \vartheta_1 < \pi$, so that $|\eta'| \geqslant 1$ in $\langle\vartheta_0, \vartheta_1\rangle$. Then

$$(d) \qquad \left|\int_{\vartheta_0}^{\vartheta_1} \eta^{an}(\vartheta)\cos n\vartheta\, d\vartheta\right| \leqslant -\int_{\vartheta_0}^{\vartheta_1} \eta^{an}(\vartheta)\eta'(\vartheta)\, d\vartheta = \frac{\eta^{an+1}}{an+1}\Big]_{\vartheta_1}^{\vartheta_0}$$

$$\leqslant \frac{1}{an+1} = O\left(\frac{1}{n}\right),$$

uniformly for $a \geqslant a_0 > 0$. Finally,

$$(e) \qquad \left|\int_{\vartheta_1}^{\pi} \eta^{an}(\vartheta)\cos n\vartheta\, d\vartheta\right| \leqslant \int_{\vartheta_1}^{\pi} \eta^{an}(\vartheta)\, d\vartheta \leqslant \eta^{an}(\vartheta_1)\,\pi = O\left(\frac{1}{n}\right).$$

3.5. Now let $0 \leqslant \vartheta \leqslant \vartheta_0 < \pi$. We have $\Phi(e^{i\theta}) = e^{i\chi(\vartheta)}$, where

$$(a) \qquad K(e^{i\chi}) = \frac{1}{\rho} K(e^{i\vartheta}); \quad \sin \tfrac{1}{2}\chi = \sqrt{\rho} \sin \tfrac{1}{2}\vartheta.$$

Hence $\chi > \vartheta$ if $\vartheta > 0$, $\chi(0) = 0$ and, by (3.3.a), $\chi(\vartheta_0) = \pi$. Further,

$$(b) \qquad \cos \tfrac{1}{2}\chi \cdot \chi' = \sqrt{\rho} \cos \tfrac{1}{2}\vartheta.$$

Hence $\chi' > \sqrt{\rho}$ if $\vartheta > 0$, and $\chi'(0) = \sqrt{\rho}$, $\chi'(\vartheta_0) = +\infty$. Finally,

$$(c) \qquad \cos \tfrac{1}{2}\chi \cdot \chi'' - \tfrac{1}{2} \sin \tfrac{1}{2}\chi \cdot \chi'^2 = -\tfrac{1}{2}\sqrt{\rho} \sin \tfrac{1}{2}\vartheta = -\tfrac{1}{2} \sin \tfrac{1}{2}\chi,$$

$$\cos \tfrac{1}{2}\chi \cdot \chi'' = \tfrac{1}{2} \sin \tfrac{1}{2}\chi(\chi'^2 - 1).$$

Hence $\chi'' > 0$ if $\vartheta > 0$, and $\chi''(0) = 0$, $\chi''(\vartheta_0) = +\infty$, so that $\chi'(\vartheta)$ is an increasing function.

Now

$$(3.5.1) \qquad 2\int_0^{\vartheta_0} \cos\left(an\chi(\vartheta)\right) \cos n\vartheta \, d\vartheta$$

$$= \int_0^{\vartheta_0} \cos\left(n\psi_1(\vartheta, a)\right) d\vartheta + \int_0^{\vartheta_0} \cos\left(n\psi(\vartheta, a)\right) d\vartheta,$$

where

$$(3.5.2) \qquad \psi_1(\vartheta, a) = a\chi(\vartheta) + \vartheta; \quad \psi(\vartheta, a) = a\chi(\vartheta) - \vartheta.$$

We have

$$(d) \qquad \psi_1' = a\chi' + 1 \geqslant 1; \quad \psi_1'' = a\chi'' \geqslant 0.$$

Hence, by the corollary to Lemma 3,

$$(3.5.3) \qquad \int_0^{\vartheta_0} \cos\left(n\psi_1(\vartheta, a)\right) d\vartheta = O\left(\frac{1}{n}\right),$$

uniformly for all $a > 0$. Further,

$$(e) \qquad \psi' = a\chi' - 1; \quad \psi'' = a\chi'' \geqslant 0.$$

Let us now suppose that

$$(3.5.4) \qquad 0 < a_0 \leqslant a \leqslant a_1 < \frac{1}{\sqrt{\rho}}.$$

Then the equation $\psi'(\tau) = 0$, *i.e.*, $\chi'(\tau) = 1/a$, has exactly one root τ with $0 < \tau_a < \vartheta_0$. From (a) and (b) it follows that

$$(f) \quad \sin^2 \tfrac{1}{2}\chi + \cos^2 \tfrac{1}{2}\chi \cdot \chi'^2 = \rho; \quad \rho \sin^2 \tfrac{1}{2}\vartheta + (1 - \rho \sin^2 \tfrac{1}{2}\vartheta)\chi'^2 = \rho,$$

and hence that

$$(g) \qquad \sin^2 \tfrac{1}{2}\tau = \frac{1-\rho\alpha^2}{\rho(1-\alpha^2)}.$$

It follows that τ_α is a continuous decreasing function of α, with $\tau(1/\sqrt{\rho}) = 0$ and $\tau(0) = \vartheta_0$. The condition (3.5.4), therefore, secures that $0 < \tau_{\alpha_1} \leqslant \tau_\alpha \leqslant \tau_{\alpha_0} < \vartheta_0$, and so

$$(h) \qquad \psi''(\tau_\alpha) \geqslant d > 0, \quad |\psi'''(\tau_\alpha)| \leqslant C,$$

uniformly in $\langle \alpha_0, \alpha_1 \rangle$. Hence, by the corollary to Lemma 4,

$$(3.5.5) \quad \int_0^{\vartheta_0} \cos\left(n\psi(\vartheta, \alpha)\right) d\vartheta = \sqrt{\left(\frac{2}{n\psi''(\tau_\alpha)}\right)}\left[\cos\left(n\psi(\tau_\alpha)\right)\int_0^\infty \frac{\cos u}{\sqrt{u}}\,du - \sin\left(n\psi(\tau_\alpha)\right)\int_0^\infty \frac{\sin u}{\sqrt{u}}\,du\right] + O(n^{-\frac{3}{2}}),$$

uniformly for these values of α.

3.6. Collecting the above results we have:

If $\rho > 1$ is given, then

$$(3.6.1) \quad a_n^{(k)} = \frac{1}{\pi}\sqrt{\left(\frac{2}{n\psi''(\tau_\alpha)}\right)}\left[\cos\left(n\psi(\tau_\alpha)\right)\int_0^\infty \frac{\cos u}{\sqrt{u}}\,du - \sin\left(n\psi(\tau_\alpha)\right)\int_0^\infty \frac{\sin u}{\sqrt{u}}\,du\right] + O(n^{-\frac{3}{2}}),$$

uniformly for all $k = an$ with $0 < \alpha_0 \leqslant \alpha \leqslant \alpha_1 < 1/\sqrt{\rho}$. Here

$$(3.6.2) \quad \psi(\vartheta) = \alpha\chi(\vartheta) - \vartheta; \quad \sin\tfrac{1}{2}\chi = \sqrt{\rho}\sin\tfrac{1}{2}\vartheta; \quad \sin^2\tfrac{1}{2}\tau_\alpha = \frac{1-\rho\alpha^2}{\rho(1-\alpha^2)}.$$

3.7. Our last step in preparing the proof of Theorem IV is the following estimate:

If $0 < \alpha_0 \leqslant \alpha = k/n \leqslant \alpha_1 < 1/\sqrt{\rho}$, and if n is sufficiently large, then, for at least one of two such consecutive integers k, we have

$$(3.7.1) \qquad |a_n^{(k)}| \geqslant \frac{D}{\sqrt{n}},$$

where $D > 0$ is independent of k and n.

Proof. We have $\chi'(\tau_\alpha) = 1/\alpha$. Hence, since $\psi(\tau_\alpha) = \alpha\chi(\tau_\alpha) - \tau_\alpha$,

$$(a) \qquad \frac{\partial\psi(\tau_\alpha)}{\partial\alpha} = \chi(\tau_\alpha) + \alpha\chi'(\tau_\alpha)\frac{\partial\tau_\alpha}{\partial\alpha} - \frac{\partial\tau_\alpha}{\partial\alpha} = \chi(\tau_\alpha).$$

Now, if n is sufficiently large,

$$(b) \quad \psi(\tau_{\alpha+1/n}) - \psi(\tau_\alpha) = \frac{1}{n}\frac{\partial\psi(\tau_\alpha)}{\partial\alpha} + \frac{1}{2n^2}\frac{\partial^2\psi(\tau_\alpha)}{\partial\alpha^2}\Big|_{\alpha=\alpha^*} \qquad (\alpha < \alpha^* < \alpha + 1/n).$$

Here

$$(c) \qquad \frac{\partial^2\psi(\tau_\alpha)}{\partial\alpha^2} = \frac{\partial}{\partial\alpha}\chi(\tau_\alpha) = \chi'(\tau_\alpha)\frac{\partial\tau_\alpha}{\partial\alpha} = \frac{1}{\alpha}\frac{\partial\tau_\alpha}{\partial\alpha} = O(1),$$

uniformly in $\langle a_0, a_1\rangle$, since $0 < \tau_{a_1} \leqslant \tau_\alpha \leqslant \tau_{a_0} < \vartheta_0$.

Hence we get from (a) and (b)

$$(d) \qquad \psi(\tau_{\alpha+1/n}) - \psi(\tau_\alpha) = \frac{1}{n}\chi(\tau_\alpha) + O\left(\frac{1}{n^2}\right),$$

uniformly in $\langle a_0, a_1\rangle$. Since $0 < \chi(\tau_{a_1}) \leqslant \chi(\tau_\alpha) \leqslant \chi(\tau_{a_0}) < \pi$, it follows that, for large n,

$$(e) \qquad 0 < \beta \leqslant n[\psi(\tau_{\alpha+1/n}) - \psi(\tau_\alpha)] \leqslant \gamma < \pi.$$

Now the roots of any equation of the form

$$(f) \qquad \cos(n\psi)A - \sin(n\psi)B = 0 \quad (A^2 + B^2 > 0),$$

are at distance π/n apart. Therefore (e) implies that there is a constant $C \geqslant 0$, not depending on α and n, such that for large values of n

$$(g) \quad \left|\cos\left(n\psi(\tau_\alpha)\right)\int_0^\infty \frac{\cos u}{\sqrt{u}}\,du - \sin\left(n\psi(\tau_\alpha)\right)\int_0^\infty \frac{\sin u}{\sqrt{u}}\,du\right| \geqslant C$$

for at least one value of each pair α, $\alpha + 1/n$, that is of each pair $k = n\alpha$, $k+1 = n(\alpha+1/n)$. Hence, since $\psi''(\tau_\alpha) \geqslant d > 0$, by $(3.5.h)$, we obtain the required result from $(3.6.1)$.

3.8. *Proof of Theorem IV.* Take $1 < \sqrt{\rho} < \frac{3}{2}$, $\phi(z) = \Phi_\rho(z)$ and $\frac{1}{2} \leqslant \alpha = k/n \leqslant \frac{2}{3}$. Then, by $(1.7.2)$ and $(3.7.1)$, if n is large enough, we have

$$(3.8.1) \qquad K_n^* \geqslant \sum_{k=\frac{1}{2}n}^{\frac{3}{4}n} |a_n^{(k)}| \geqslant \tfrac{1}{2}\cdot\tfrac{1}{6}n\cdot\frac{D}{\sqrt{n}} = \frac{D}{12}\sqrt{n}.$$

3.9. *Proof of Theorem V.* Take $1 < \sqrt{\rho} < \frac{3}{2}$, $\phi(z) = \Phi_\rho(z)$ and define the polynomial $F_n(z)$ of degree n by

(3.9.1) $A_k^{(n)} = \operatorname{sign} a_n^{(k)}$, if $\frac{1}{2}n \leqslant k \leqslant \frac{2}{3}n$; $A_k^{(n)} = 0$ elsewhere.

Then for the function $f_n(z) = F_n\big(\phi(z)\big) = \sum_1^\infty a_m^{(n)} z^m$, since

$$a_m^{(n)} = \sum_{k=1}^{m} A_k^{(n)} a_m^{(k)},$$

we have

(3.9.2) $|a_m^{(n)}| \leqslant \frac{2}{3}n$ for all m; $a_m^{(n)} = 0$, if $m < \frac{1}{2}n$.

Moreover, by (3.8.1),

(3.9.3) $$a_n^{(n)} = \sum_{k=\frac{1}{2}n}^{\frac{2}{3}n} |a_n^{(k)}| \geqslant \frac{D}{12}\sqrt{n}.$$

Now consider

(3.9.4) $$F(z) = \sum_{p=1}^{\infty} F_{n_p}(z); \quad f(z) = \sum_{p=1}^{\infty} f_{n_p}(z) = F\big(\phi(z)\big),$$

where

(3.9.5) $$n_{p+1} > 2n_p, \quad \sum_{q<p} n_q = o(\sqrt{n_p}).$$

The modulus of every coefficient A_k of $F(z)$ is either one or zero, and $f(z)$ is schlicht subordinate to $F(z)$; but

(3.9.6) $$a_n \neq o(\sqrt{n}).$$

For consider the coefficient a_{n_p}. The contribution of $f_{n_p}(z)$ to this coefficient is, by (3.9.3),

(*a*) $$a_{n_p}^{(n_p)} \geqslant \frac{D}{12}\sqrt{n_p}.$$

The contribution of all the $f_{n_q}(z)$ with $q < p$ is, by (3.9.2) and (3.9.5), of order

(*b*) $$O\Big(\sum_{q<p} n_q\Big) = o(\sqrt{n_p});$$

while, again by (3.9.2) and (3.9.5), the $f_{n_q}(z)$ with $q > p$ do not contribute at all to a_{n_p}.

In a similar way we can obviously construct a function $F(z)$ with $A_n = O(n^\alpha)$, $\alpha > 0$, and a function $f(z) \prec F(z)$ with $a_n \neq o(n^{\alpha+\frac{1}{2}})$.

Bibliography.

Bernstein, S., "Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degré donné", *Mémoire couronné par la classe des sciences de l'Académie royale de Belgique*, 1912, 19–20.

Bieberbach, E., *Lehrbuch der Funktionentheorie*, 2 (Berlin, 1927).

Carathéodory, C., "Über den Variabilitätsbereich der Koefficienten von Potenzreihen, die gegebene Werte nicht annehmen", *Math. Annalen*, 64 (1907), 95–115.

Dieudonné, J., "Recherches sur quelques problèmes relatifs aux polynomes et aux fonctions bornées d'une variable complexe", *Annales Sci. de l'Ecole Normale* (3), 48 (1931), 247–358.

Fejér, L., "Über die Positivität von Summen, die nach trigonometrischen oder Legendreschen Funktionen fortschreiten", *Acta Litt. ac Scient. Szeged* (1925), 75–86.

Koebe, P., "Über das Schwarzsche Lemma und einige damit zusammenhängende Ungleichheitsbeziehungen der Potentialtheorie und Funktionentheorie", *Math. Zeitschrift*, 6 (1920), 52–84.

Landau, E., *Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie* (Ed. 2, Berlin, 1929).

Lindelöf, E., "Mémoire sur certaines inégalités dans la théorie des fonctions monogènes et quelques propriétés nouvelles de ces fonctions dans le voisinage d'un point singulier essentiel", *Acta Soc. Sci. Fennicae*, 35 (1909), 1–35.

Littlewood, J. E., "On inequalities in the theory of functions", *Proc. London Math. Soc.* (2), 23 (1924), 481–519.

Löwner, K., "Untersuchungen über schlichte konforme Abbildungen des Einheitskreises", *Math. Annalen*, 89 (1923), 103–121.

Pólya-Szegö, *Aufgaben und Lehrsätze aus der Analysis*, 1 & 2 (Berlin, 1925).

Prawitz, H., "Über Mittelwerte analytischer Funktionen", *Ark. for Math. Astron. och Fys*, 20A (1927), 1–12.

Riesz, M., "Eine trigonometrische Interpolationsformel und einige Ungleichungen für Polynome", *Jahresber. der Deutsch. Math. Ver.*, 23 (1915), 354–368.

Rogosinski, W., **1**, "Über Bildschranken bei Potenzreihen und ihren Abschnitten", *Math. Zeitschrift*, 17 (1923), 260–276.

————, **2**, "Über positive harmonische Entwicklungen und typisch-reelle Potenzreihen", *Math. Zeitschrift*, 35 (1932), 93–121.

————, **3**, "Über den Wertevorrat einer analytischen Funktion", *Schrift. Königsberg. Gel. Ges., Naturw. Klasse*, 8 (1931), 1–31.

————, **4**, "Zum Majorantenprinzip der Funktionentheorie", *Math. Zeitschrift*, 37 (1933), 210–236.

————, **5**, "On subordinate functions", *Proc. Camb. Phil. Soc.*, 35 (1939), 1–26.

Schur, I., "Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind", *Journal für Math.*, 147 (1917), 205–232; 148 (1918), 122–145.

Spencer, D. C., "On finitely mean valent functions", *Proc. London Math. Soc.* (2), 47 (1941), 201–211.

GROUPOIDS AND THEIR AUTOMORPHISMS

By A. R. RICHARDSON.

[Received 23 April, 1940.—Read 9 May, 1940.]

The mathematical domain is greatly restricted when the associative and commutative laws are postulated. Thus of the finite number systems of order n closed to a binary operation, about n^{-n} of them are associative, $n^{-n(n-1)/2}$ are commutative and $n^{-(2n-1)}$ have a unit or modulus. However, many well-known theorems, *e.g.* the Jordan-Hölder theorem, proved for associative systems, are of much wider extent; some of these will be discussed here as well as certain theorems which have only trivial analogues in associative algebra. Restrictive postulates will be introduced as required so that the later theorems of this paper are less likely to be satisfied in an arbitrary mathematical domain than the earlier ones. Thus Theorems 1–17 (except 2) are true for *any set* $\mathcal{G}$ *and for any fixed group of correspondences* although, for the sake of continuity, $\mathcal{G}$ is taken to be a groupoid and the group of correspondences is referred to as the group of automorphisms of $\mathcal{G}$. In the remainder of the work it is essential that $\mathcal{G}$ shall be a groupoid and that the correspondences shall be automorphisms of $\mathcal{G}$. In some theorems it is necessary to postulate certain weak forms of the associative or commutative laws. The final theorems relate to inner automorphisms, some of which have only trivial analogues in associative sets.

Definition 1. The basic system will be termed a *Groupoid*. It is a system, of finite or infinite order, closed to a binary operation termed *multiplication*, *i.e.*, if a and b are any two elements of the system, then so also is $a.b$.

A definition of *equality* ensures that equals may be substituted for equals, *i.e.*, if $a = b$ then $ac = bc$ and $ca = cb$.

These groupoids form only an infinitesimal part of the set of all multiplicative systems, which includes sets not closed to the operation and also sets in which the product $a.b$ is a set of elements, *e.g.* vectors, and not necessarily a single element as in a groupoid.

Notation. $\mathcal{G}$ denotes the basic groupoid, $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, ... denote sub-groupoids of $\mathcal{G}$; $a, b, c, \ldots$ denote elements of $\mathcal{G}$; and Σ denotes a set of elements of $\mathcal{G}$.

$G(\mathcal{G})$ denotes the group of automorphisms of $\mathcal{G}$; $S, T, \ldots$ denotes elements of $G(\mathcal{G})$; a^s denotes the element of $\mathcal{G}$ corresponding to a in the automorphism S.

$A, B, C, \ldots$ denote sub-groups of $G(\mathcal{G})$.

R denotes an equivalence relation over $\mathcal{G}$; $\mathfrak{R}$ denotes the set of all equivalence relations over $\mathcal{G}$.

Definition 2. a^s is termed the *conjugate* of a under s.

1. *Congruences and equivalence relations.*

Let $\mathcal{G}$ be separated into disjoint classes where, in order to avoid the paradoxes connected with inconsistent classes, the admission of an element to a class does not depend on elements previously considered.

Let $a \equiv b \pmod{R}$ denote that b is in the same class as a. Hence (i) $a \equiv a$; (ii) if $a \equiv b$, then $b \equiv a$; (iii) if $a \equiv b$ and $b \equiv c$, then $a \equiv c \pmod{R}$.

Hence R is an *equivalence relation**.

Conversely, any equivalence relation R over $\mathcal{G}$ defines a separation of the elements of $\mathcal{G}$ into disjoint classes (mod R).

THEOREM 1. *The set $\mathfrak{R}$ of all equivalence relations over $\mathcal{G}$ may be ordered.*

Let $R_1 \subset R_2 \pmod{P}$ if $a \equiv b \pmod{R_1}$ implies that $a \equiv b \pmod{R_2}$. Hence $R_1 \subset R_1 \pmod{P}$; also, if $R_1 \subset R_2 \pmod{P}$ and $R_2 \subset R_3 \pmod{P}$, then $R_1 \subset R_3 \pmod{P}$. Therefore $\mathfrak{R}$ is ordered (mod P).

$\mathfrak{R}$ may be ordered in other ways, but throughout this paper it will be assumed to be ordered (mod P).

---

* G. Birkhoff.

When this ordered $\mathfrak{R}$ is a *structure**, the cross-cut and union may be defined thus:—

Definition 3. The *cross-cut*, $D = (R_1, R_2)$, of any two equivalence relations is such that $a \equiv b \pmod{D}$ if, and only if, $a \equiv b \pmod{R_1}$ and also $a \equiv b \pmod{R_2}$.

D always exists for the *unit equivalence relation*, E, in which every element is the sole element in the class (mod E), is such that $E \subset R$.

Evidently $D \subset R_1$ and $D \subset R_2$.

Also D is maximal, *i.e.* if $C \subset R_1$ and $C \subset R_2$ then $C \subset D$.

Also D is an equivalence relation.

Definition 4. The *union*, $M = [R_1, R_2]$, is such that $a \equiv b \pmod{M}$ if, and only if, $u_1, u_2, \ldots, u_{2s-1}$ exist such that

$$a \equiv u_1 \pmod{R_1},\ u_1 \equiv u_2 \pmod{R_2},\ u_2 \equiv u_3 \pmod{R_1},$$

$$\ldots,\ u_{2s-1} \equiv b \pmod{R_2}.$$

Evidently $R_1 \subset M$ and $R_2 \subset M$.

Also M is minimal, *i.e.* if $R_1 \subset C$ and $R_2 \subset C$ then $M \subset C$, for, if $a \equiv b \pmod{M}$, then $u_1, u_2, \ldots, u_{2s-1}$ exist such that

$$a \equiv u_1 \pmod{R_1},\ u_1 \equiv u_2 \pmod{R_2},\ \ldots,\ u_{2s-1} \equiv b \pmod{R_2}$$

and, since $R_1 \subset C$, $R_2 \subset C$,

$$a \equiv u_1 \pmod{C},\ u_1 \equiv u_2 \pmod{C},\ \ldots,\ u_{2s-1} \equiv b \pmod{C}.$$

Hence $a \equiv b \pmod{C}$, therefore $M \subset C$.

Also M is an equivalence relation, for $a \equiv a \pmod{M}$; if $a \equiv b \pmod{M}$ then $u_1, u_2, \ldots, u_{2s-1}$ exist such that $b \equiv u_{2s-1} \pmod{R_1}$, $u_{2s-1} \equiv u_{2s-2} \pmod{R_1}$, $\ldots$, $u_2 \equiv u_1 \pmod{R_2}$, $u_1 \equiv a \pmod{R_2}$, *i.e.* $b \equiv a \pmod{M}$; if $a \equiv b \pmod{M}$ and $b \equiv c \pmod{M}$, then $u_1, u_2, \ldots, u_{2s-1}$; $v_1, v_2, \ldots, v_{2t-1}$ exist such that

$$a \equiv u_1 \pmod{R_1},\ u_1 \equiv u_2 \pmod{R_2},\ \ldots,\ u_{2s-1} \equiv b \pmod{R_2},$$

$$b \equiv v_1 \pmod{R_1},\ \ldots,\ v_{2t-1} \equiv c \pmod{R_2};$$

hence $a \equiv c \pmod{M}$. Therefore M is an equivalence relation.

---

* O. Ore, *Annals of Math.*, 36 (1935), 408.

Hence the set $\mathfrak{R}$ of all equivalence relations over $\mathcal{G}$ is a structure. It is not a Dedekind structure* but contains a subset which is Dedekindian†.

THEOREM 2. *The set of all automorphisms of $\mathcal{G}$ is a group.*

The set contains an identical automorphism E or I in which $a \rightleftarrows a$, for $a^{ES} = (a^E)^S = a^S = (a^S)^E = a^{SE}$ for all a, *i.e.* $ES = S = SE$.

If T and S are automorphisms in $T : a \to a^T$; in $S : a^T \to (a^T)^S$; hence in $TS : a \to (a^T)^S$, *i.e.* $(a^T)^S = a^{TS}$. Similarly, $b \to (b^T)^S = b^{TS}$. Since T is an automorphism $a \,.\, b \to a^T b^T = (a \,.\, b)^T$ and, since S is an automorphism, $a^T \,.\, b^T \to (a^T \,.\, b^T)^S = (a^T)^S (b^T)^S = a^{TS} b^{TS}$. Hence in TS $a \,.\, b \to a^{TS} \,.\, b^{TS}$, *i.e.* TS is an automorphism.

Also every automorphism has an inverse; if in $T : a \to a^T$, then a (1, 1) correspondence T' exists in which $T' : a^T \to a$ and $TT' = E = T'T$, *i.e.* $T' = T^{-1}$ and T' is an automorphism.

Finally, $(RS)T = R(ST)$ for in $RS : a \to (a^R)^S$ and in $(RS)\,T\; a \to [(a^R)^S]^T$. In $ST : a^R \to [a^R)^S]^T$, hence in $R(ST)\; a \to [(a^R)^S]^T$, *i.e.* $(RS)\,T = R(ST)$.

Therefore the set of all automorphisms is a group.

This group is evidently a sub-group of the group of all (1, 1) correspondences which is itself a sub-group of the groupoid of all correspondences over $\mathcal{G}$.

The group of automorphisms $G(\mathcal{G})$ is isomorphic with the permutation group, $\begin{pmatrix} a \\ a^S \end{pmatrix}$, on the elements of $\mathcal{G}$. In such cases $\mathcal{G}$ will be termed a *transitive*, or *primitive*, or *regular* groupoid according as $G(\mathcal{G})$ has the corresponding properties. Hence the sets of elements of $\mathcal{G}$ which are conjugates under $G(\mathcal{G})$ are the *transitive sets* of $G(\mathcal{G})$ when this is exhibited as a permutation group on the elements of $\mathcal{G}$.

The choice of terms in which to describe groupoids possessing properties not peculiar to those of permutation groups is more difficult, *e g.* the term *cyclic* may be taken to refer either to the property that every element is a power of some fixed element or that $G(\mathcal{G})$ is cyclic. Such questions will not be discussed here, but mention may be made of the theorem that a finite, transitive groupoid for which $G(\mathcal{G})$ is cyclic is generated from a single element and is therefore cyclic in both senses.

THEOREM 3. *The set of all automorphisms S for which $a \equiv a^S \pmod{R}$ for all a is a group.*

---

* O. Ore, *Annals of Math.*, 36 (1935), 412.

† P. Dubreil et M.-L. Dubriel-Jacobin, *Journal de Math.* (9), 18 (1939), 81.

If S, T belong to the set, then $a \equiv a^S$, $a^S \equiv (a^S)^T = a^{ST} \pmod{R}$. Hence $a \equiv a^{ST} \pmod{R}$; therefore ST also belongs to the set. Evidently the unit automorphism E belongs to the set and, since $(a^{S^{-1}})^S = a$, S^{-1} also belongs. Therefore the set is a group.

Definition 5. This group will be termed the *group of inertia* and denoted by G_R.

It is a sub-group of $G(\mathcal{G})$ and is the set of all elements of $G(\mathcal{G})$ which interchanges the elements of each class (mod R) amongst themselves. It is therefore the maximal sub-group having this property.

Examples of inertial groups are:—

(*a*) The group which leaves a given element a invariant. This group will be denoted by C_a. If $\mathcal{G}$ is regular, $C_a = 1$.

(*b*) The group which *centralizes* a sub-groupoid $\mathfrak{H}$ or a sub-set Σ, *i.e.* which leaves every element of $\mathfrak{H}$ or Σ invariant. It will be denoted by $G_{\mathfrak{H}}$ or G_Σ.

(*c*) The *group of decomposition* of $\mathfrak{H}$, viz. the set of all automorphisms which interchange the elements of $\mathfrak{H}$, *i.e.* which leave $\mathfrak{H}$ invariant. This group will be denoted by $G_s(\mathfrak{H})$.

These groups generalize the correspondingly-named groups in the theory of algebraic corps.

Since $G(\mathcal{G})$ may be of more than countable order it is necessary to define:—

(i) The *cross-cut* of any aggregate of sub-sets of $G(\mathcal{G})$ is the set of elements common to all the sets of the aggregate.

If the sub-sets are sub-groups, the cross-cut is also a sub-group. Hence $D = (A, B)$ is the set of all elements common to A and B.

(ii) The *union*. Let Σ be any set of elements of $G(\mathcal{G})$. Let Θ denote the set of all sub-groups of $G(\mathcal{G})$ which contain Σ as a sub-set. The union, M, of Σ is the cross-cut of Θ, *i.e.*, M is the minimal sub-group which contains all the elements of Σ.

THEOREM 4. *The union of A and B is the set M of all finite products $a_1 b_1 \dots a_s b_s$ of elements of A and B.*

M is a group, for if m_1, m_2 are elements, so is $m_1 m_2$; it contains the unit element, and, if $m_1 = a_1 b_1 \dots a_s b_s$, then $m_1^{-1} = b_s^{-1} a_s^{-1} \dots b_1^{-1} a_1^{-1}$. Hence M is a sub-group of $G(\mathcal{G})$ which contains all the elements of A and B,

and, since it contains no elements which are not products of elements of A and B, it is minimal. Therefore $M = [A, B]$.

Two results follow at once from the definitions:—

(i) $\left(G_z(\mathfrak{B}), G_z(\mathfrak{C})\right) \leqslant G_z(\mathfrak{B}, \mathfrak{C})$.

(ii) If G_i denotes $G_z(\Sigma_i)$ where Σ_i are the classes of $\mathfrak{Q}$ (mod R), then $(G_1, G_2 \ldots, G_{i-1}, G_{i+1}, \ldots) \leqslant G_i$.

THEOREM 5. *$G_{\mathfrak{B}}$ is a normal divisor of $G_z(\mathfrak{B})$.*

Let a be any element in $\mathfrak{B}$, S in $G_{\mathfrak{B}}$, z in $G_z(\mathfrak{B})$; then $a^{z^{-1}}$ is in $\mathfrak{B}$; hence $(a^{z^{-1}})^S = a^{z^{-1}S} = a^{z^{-1}}$ and $(a^{z^{-1}})^z = a$. Hence $z^{-1}Sz$ leaves every element of $\mathfrak{B}$ invariant and is therefore in $G_{\mathfrak{B}}$, *i.e.*, $G_{\mathfrak{B}}$ is a normal divisor of $G_z(\mathfrak{B})$. (See also Theorem 19.)

THEOREM 6. *The set of all automorphisms S which map a fixed element a on a fixed element b is $C_a S_1$, where S_1 is a fixed automorphism for which $b = a^{S_1}$.*

If $b = a^T = a^{S_1}$ then $T = CS_1$, where C is in C_a. Also if C_2 is any element of C_a, then $T_2 = C_2 S_1$ also belongs to the set, for

$$a^{T_2} = a^{C_2 S_1} = (a^{C_2})^{S_1} = a^{S_1} = b.$$

Hence T_2 belongs to the coset $C_a S_1$ and every element of this coset maps a on $a^{S_1} = b$.

COROLLARY 1. $S_1^{-1} C_a S_1 = C_b$, *i.e.* $C_a \cong C_b$ *if* $a \equiv b \bmod G(\mathfrak{Q})$.

COROLLARY 1. *There is a* (1, 1) *correspondence between the complete set of all the conjugates of a under $G(\mathfrak{Q})$ and the cosets of*

$$G(\mathfrak{Q}) = C_a + \ldots + Ca\, S_i + \ldots,$$

where S_i maps a on $a^{S_i} = a_i$, viz., $a_i \rightleftarrows C_a S_i$.

COROLLARY 2. *Since the elements in any two cosets $Ca\, S_1$, $Ca\, S_2$ can be put into* (1, 1) *correspondence, the number of conjugates of a under $G(\mathfrak{Q})$ is a divisor of the order of $G(\mathfrak{Q})$. Further, if $a \equiv b \equiv c$ [mod $G(\mathfrak{Q})$] then the number of automorphisms which map a on b is the same as the number which map a on c or b on c.*

This corollary is significant only when the order of $G(\mathfrak{Q})$ is finite. It then generalizes the group theorem that the order of a class of conjugate elements divides the group order.

THEOREM 7. *If* $R_1 \subset R_2$ *then* $G_{R_1} \leqslant G_{R_2}$.

Let S be any automorphism in G_{R_1}; then, for all a in $\mathfrak{G}$, $a \equiv a^S \pmod{R_1}$. Therefore $a \equiv a^S \pmod{R_2}$, *i.e.* S is in G_{R_2}. Hence $G_{R_1} \leqslant G_{R_2}$.

THEOREM 8. $(G_{R_1}, G_{R_2}) = G_{(R_1, R_2)}$.

Let a be any element in $\mathfrak{G}$ and let S be any automorphism in (G_{R_1}, G_{R_2}). Hence $a \equiv a^S \pmod{R_1}$ and $a \equiv a^S \pmod{R_2}$. Therefore $a \equiv a^S$ [mod (R_1, R_2)], *i.e.* S is in $G_{(R_1, R_2)}$. Therefore $(G_{R_1}, G_{R_2}) \leqslant G_{(R_1, R_2)}$.

Next, let S be any automorphism in $G_{(R_1, R_2)}$; then, for all a in $\mathfrak{G}$, $a \equiv a^S$ [mod (R_1, R_2)], *i.e.* $a \equiv a^S \pmod{R_1}$ and $a \equiv a^S \pmod{R_2}$. Therefore S is in both G_{R_1} and G_{R_2}, *i.e.* is in (G_{R_1}, G_{R_2}). Hence

$$G_{(R_1, R_2)} \leqslant (G_{R_1}, G_{R_2}), \text{ i.e. } (G_{R_1}, G_{R_2}) = G_{(R_1, R_2)}.$$

THEOREM 9. $[G_{R_1}, G_{R_2}] \leqslant G_{[R_1, R_2]}$.

Let M be any element of $[G_{R_1}, G_{R_2}]$, *e.g.* $M = S_1 T_1 S_2 T_2 \ldots S_n T_n$, where $S_1, S_2, \ldots, S_n$ are in G_{R_1} and $T_1, T_2, \ldots, T_n$ are in G_{R_2}. Let a be any element in $\mathfrak{G}$; then

$$a \equiv a^{S_1} \pmod{R_1}, \quad a^{S_1} \equiv a^{S_1 T_1} \pmod{R_2}, \quad \ldots, \quad a^{S_1 T_1 S_2 T_2 \ldots S_n} \equiv a^M \pmod{R_2}.$$

Therefore $$a \equiv a^M \pmod{[R_1, R_2]},$$

i.e. M is in $G_{[R_1, R_2]}$. Hence

$$[G_{R_1}, G_{R_2}] \leqslant G_{[R_1, R_2]}$$

In general, equality does not occur.

Definition 6. Equivalence relations which give rise to the same inertial group will be termed *equi-inertial* or *isotopic*.

COROLLARY. *If* $R_1 \subset R_2$ *then* $[R_1, (R_2, R_3)] \subset (R_2, [R_1, R_3])$. *Hence*

$$[G_{R_1}, G_{(R_2, R_3)}] \leqslant G_{[R_1, (R_2, R_3)]} \leqslant G_{(R_2, [R_1, R_3])} \leqslant (G_{R_2}, G_{[R_1, R_3]}),$$

i.e. $$[G_{R_1}, (G_{R_2}, G_{R_3})] \leqslant (G_{R_2}, G_{[R_1, R_3]}).$$

Let Σ be any class of elements of $\mathfrak{G}$ (mod R) and let a be any element in Σ. Then the complete set of conjugates of a under G_R, *i.e.* the set of elements a^S, where S ranges over G_R, is in Σ.

Assume therefore that Σ can be separated into disjoint sets, each consisting exclusively of elements conjugate under G_R. This will certainly

be the case if $\mathfrak{Q}$ can be well-ordered. Let R_2 denote an equivalence relation obtained by separating some of these disjoint classes. Then $R_2 \subset R$ and $G_{R_2} \leqslant G_R$. But G_R leaves every class in R_2 invariant, therefore $G_R \leqslant G_{R_2}$, *i.e.* $G_{R_2} = G_R$.

It follows that an equivalence relation R_m exists in which each class consists exclusively of a single complete set of conjugates under G_R and that R_m is *minimal* for the isotopic equivalence relations having G_R for inertial group.

Although R_m exists, it does not follow that the descending chain condition holds.

A few properties of isotopic chains will now be mentioned. Let R_m be minimal and denote its classes by $\bar{a}$, $\bar{b}$, $\bar{c}$, ...; then, if no automorphism exists which maps an element of $\bar{a}$ on an element of $\bar{b}$, the equivalence relation R_2 in which the classes are $\{\bar{a}, \bar{b}\}$, $\bar{c}$, ... has G_{R_m} for its inertial group. Also*, R_2 is prime over R; for, if $R_m \subset R_1 \subset R_2$, then $G_{R_1} = G_{R_m}$. But, since $R_m \subset R_1$, each class $\bar{a}$, $\bar{b}$, $\bar{c}$, ... of R_m is in some class of R_1 and each class of R_1 is in some class of R_2; and, since only one class in R_2 differs from a class in R_m, either $R_m = R_1$ or $R_1 = R_2$, *i.e.* R_2 is prime over R_m.

Similarly, R_{i+1} is prime over R_i if exactly one class of R_{i+1} consists of two, and only two, classes of R_i.

By Theorem 8, if R_1, R_2 are in different chains, then (R_1, R_2) is in both chains; therefore R_m is in all descending isotopic chains and the theory of quotients and transformations due to O. Ore† is applicable.

In general, ascending isotopic chains do not have a unique *maximal* R_m.

Let R_3 be derived from R_2 by removing one of the elements, *e.g.*, a, from a class (mod R_2), leaving the other classes unaltered. Then, R_2 being minimal, R_2 is prime over R_3 and $G_{R_3} = (C_a, G_{R_2})$.

THEOREM 10. *If R is minimal, if $R \subset R_i$, if G_R is a normal divisor of G_{R_i} and if $G_{R_i} = \Sigma G_R S$, then every element of the* coset $G_R S$ *permutes the classes* $\bar{a}$ (mod R) *among themselves in the same way.*

Since $R \subset R_i$, $G_R \leqslant G_{R_i}$. Let S be not in G_R. Hence some element a of $\mathfrak{Q}$ exists such that $a \not\equiv a^S = b$ (mod R). Since R is minimal if a^1 is any conjugate of a (mod R), an automorphism T can be found in G_R such that $a^1 = a^T$. Hence $a^{1S} = a^{TS} = a^{S \cdot S^{-1}TS} = b^{S^{-1}TS}$. Since G_R is normal in G_{R_i}, $S^{-1}TS = T_2$ is in G_R. Therefore $b^{S^{-1}TS} = b$ (mod R), *i.e.*

---

* O. Ore, *Annals of Math.*, 36 (1935), 411.

† O. Ore, *loc. cit.*, 427.

every element in $\bar{a}$ is mapped into an element in the same class $\bar{b}$ by S and therefore by every element in the coset $G_R S$.

If G_R is *not* normal in G_{R_i}, then S and T exist in G_{R_i} and G_R respectively such that $S^{-1}TS$ is not in G_R. Therefore some element a exists such that $a^{S^{-1}TS} \not\equiv a \pmod{R}$. Hence the classes (mod R) are split by the automorphisms of the conjugate sub-groups $S^{-1}G_2 S$, which, however, leave the classes (mod R_i) invariant.

Different cosets $G_R S$, $G_R T$ may map $\bar{a}$ into the same class b; for let a be any element in $\bar{a}$, P any automorphism in G_R; and then, if $a^{-S_1} = a^{-T_2}$, it must be possible to find ϕ in G_R such that $a^{PS_2} = a^{QT_2}$, *i.e.* $PST^{-1}\phi^{-1}$ is in C_a. If, therefore, it is possible to find a such that C_a is in G_{R_i} but not in G_R, then $G_R S$, $G_R T$ both map $\bar{a}$ on b, and in general this is so.

On the other hand, $G_R S$ and $G_R T$ do not map all classes in the same way; for, if a^S, a^T are in the same class, then $a^{ST^{-1}}$ is in the same class as a, *i.e.* $a^{ST^{-1}} \equiv a \pmod{R}$ for all a. Hence ST^{-1} is in G_R, *i.e.* S is in the coset $G_R T$ contrary to hypothesis.

COROLLARY 1. *The classes which are interchanged by the elements of the coset $G_R S$ lie in the same class* (mod R_i).

COROLLARY 2. *G_{R_i}/G_R can be represented as a permutation group on the symbols denoting the classes $\bar{a}$* (mod R).

COROLLARY 3. *In cases in which $\mathfrak{Q}$ contains only a countable set of elements $G(\mathfrak{Q})$ can be represented as a permutation group on symbols denoting the elements of $\mathfrak{Q}$. The classes* (mod R), *when R is minimal, are then the sets of transivity of $G(\mathfrak{Q})$ and the theory of permutation groups applies.*

THEOREM 11. *If R_1 and R_2 are minimal, then so is $[R_1, R_2]$.*

Let $a \equiv b \pmod{[R_1, R_2]}$. Then $u_1, u_2, \ldots, u_{2s-1}$ exist such that $a \equiv u_1 \pmod{R_1}$, $u_1 \equiv u_2 \pmod{R_2}$, $\ldots$, $u_{2s-1} \equiv b \pmod{R_2}$ and, since R_1, R_2 are minimal, $S_1, S_2, \ldots, S_s$; $T_1, T_2, \ldots, T_s$ exist such that

$$u_1 \equiv a^{S_1}, \quad u_2 = u_1^{T_1} = a^{S_1 T_1}, \quad \ldots, \quad a^{S_1 T_1 S_2 T_2 \ldots S_{n-1}} \equiv b \pmod{R_2},$$

i.e. $b = a^{S_1 T_1 S_2 T_2 \ldots S_{n-1} T_{n-1}}$. But, by Theorem 9, $[G_{R_1}, G_{R_2}] \leqslant G_{[R_1, R_2]}$, and therefore $S_1 T_1 S_2 T_2 \ldots S_{n-1} T_{n-1}$, which is in $[G_{R_1}, G_{R_2}]$, is also in $G_{[R_1 R_2]}$. Hence b is an image of a (mod $[R_1, R_2]$), *i.e.* $[R_1, R_2]$ is minimal.

In general, (R_1, R_2) is not minimal; for $a \equiv b \pmod{(R_1, R_2)}$ implies $a \equiv b \pmod{R_1}$ and $a \equiv b \pmod{R_2}$. Hence S, T exist in G_{R_1}, G_{R_2} respec-

tively such that $a^S = a^T = b$, *i.e.* $S = C_a T$, but we cannot assert that $S = T$. For example

$\mathcal{G}$		a	b	c	d	e
	a	b	a	b	b	b
	b	b	a	a	a	a
	c	d	e	c	a	b
	d	e	c	b	d	a
	e	e	d	a	b	e

$$G(\mathcal{G}) = 1,\ (ab)(cd),\ (ab)(de),\ (ab)(ce),\ (cde)(ced).$$

$$R_1 : \binom{a}{b}, \binom{c}{d}, e; \quad R_2 : \binom{a}{b}, \binom{d}{e}, c; \quad (R_1, R_2) : \binom{a}{b}, c, d, e;$$

$$G_{R_1} = 1,\ (ab)(cd); \quad G_{R_2} = 1,\ (ab)(de); \quad G_{(R_1, R_2)} = 1 = (G_{R_1}, G_{R_2}).$$

COROLLARY. *If $\mathcal{G}$ is regular, then $C_a = 1$ for all a, i.e. (R_1, R_2) is minimal.*

THEOREM 12. *If R is minimal and if $G_R \leqslant G_{R_1}$, then $R \subset R_1$.*

Let a be any element in $\mathcal{G}$ and let $a \equiv b \pmod{R}$. Then, since R is minimal, an automorphism S exists in G_R such that $b = a^S$. Hence $a \equiv a^S \pmod{R}$ and, since $G_R \leqslant G_{R_1}$, S also belongs to G_{R_1}. Therefore $a \equiv a^S \pmod{R_1}$, *i.e.* $a \equiv b \pmod{R}$ implies $a \equiv b \pmod{R_1}$; therefore $R \subset R_1$.

THEOREM 13. *If R is minimal the number of elements in every class (mod R) is a divisor of the order of G_R.*

By Theorem 6 the set $(C_a S_1, G_R)$ is the complete set of those automorphisms in G_R which map a on $b = a^{S_1}$. Take S_1 in G_R; the set is then $(C_a, G_R) S_1$. The number of these elements is independent of S_1 and is the order of the group (C_a, G_R). Since R is minimal, every element of the class is an image of a and is mapped on a by the same number of automorphisms in G_R. Hence, if n is the order of G_R and i is the order of (C_a, G_R), the number of elements in the class $\bar{a}$ is n/i.

THEOREM 14. *If R_1 and R_2 are minimal, $G_{[R_1, R_2]} \leqslant C_a[G_{R_1}, G_{R_2}]$.*

Since R_1 and R_2 are minimal, if a is any selected element of $\mathcal{G}$ and M is any element of $G_{[R_1, R_2]}$, then $a \equiv a^M \pmod{[R_1, R_2]}$. Hence $u_1, u_2, \ldots, u_{2s-1}$

exist such that

$$a \equiv u_1 \pmod{R_1}, \quad u_1 \equiv u_2 \pmod{R_2}, \quad \ldots, \quad u_{2s-1} \equiv a^M \pmod{R_2},$$

and $S_1, S_2, \ldots, S_{n-1}$; $T_1, T_2, \ldots, T_{n-1}$ exist in G_{R_1}; G_{R_2} respectively such that $u_1 = a^{S_1}$, $u_2 = u_1^{T_1} = a^{S_1 T_1}$, and so on, and $a^{S_1 T_1 \cdots S_n} \equiv a^M \pmod{R_2}$, *i.e.* T_n exists such that $a^M = a^{S_1 T_1 \cdots S_n T_n}$. Therefore $S_1 T_1 S_2 T_2 \ldots S_n T_n M^{-1} = C$, where C is in C_a. Keeping a fixed and allowing M to range over $G_{[R_1, R_2]}$ and noting that $S_1 T_1 \ldots S_n T_n$ belongs to $[G_{R_1}, G_{R_2}]$, we have

$$G_{[R_1, R_2]} \leqslant C_a [G_{R_1}, G_{R_2}].$$

COROLLARY 1. $G_{[R_1, R_2]} \leqslant (\ldots, C_a [G_{R_1}, G_{R_2}], \ldots)$.

COROLLARY 2. *In a regular groupoid*

$$G_{[R_1, R_2]} \leqslant [G_{R_1}, G_{R_2}],$$

i.e.

$$G_{[R_1, R_2]} = [G_{R_1}, G_{R_2}].$$

COROLLARY 3. *If the groups C_a are distributive with respect to $[G_{R_1}, G_{R_2}]$, then*

$$G_{[R_1, R_2]} \leqslant (\ldots C_a \ldots)[G_{R_1}, G_{R_2}] = [G_{R_1}, G_{R_2}],$$

since $(\ldots C_a \ldots) = 1$, *i.e.* $G_{[R_1, R_2]} = [G_{R_1}, G_{R_2}]$.

Specially related equivalence relations termed *associable* are of considerable importance.*

Definition 7. R_1 and R_2 are *associable* if, when $a \equiv u_1 \pmod{R_1}$ and $u_1 \equiv b \pmod{R_2}$, v_1 can be found such that $a \equiv v_1 \pmod{R_2}$, $v_1 = b \pmod{R_1}$.

THEOREM 15. *If G_{R_1}, G_{R_2} are permutable, then R_1, R_2 are associable if minimal.*

Let $a \equiv u_1 \pmod{R_1}$, $u_1 \equiv b \pmod{R_2}$. Then, since R_1, R_2 are minimal, S, T exist in G_{R_1}, G_{R_2} respectively such that $u_1 = a^S$, $b = u_1^T = a^{ST}$ and, since G_{R_1}, G_{R_2} are permutable, S_1, T_1 can be found such that $ST = T_1 S_1$, *i.e.* $b = a^{T_1 S_1} = (a^{T_1})^{S_1}$. Hence $a \equiv a^{T_1} \pmod{R_2}$, $a^{T_1} \equiv a^{T_1 S_1} = b \pmod{R_1}$. Therefore R_1, R_2 are associable.

In general, the converse is not true. Let R_1, R_2 be associable and minimal. Then, if $a = u \pmod{R_1}$ and $u = b \pmod{R_2}$, v exists such that $a \equiv v \pmod{R_2}$, and $v \equiv b \pmod{R_1}$, *i.e.*, S, S_1; T, T_1 exist in G_{R_1}; G_{R_2} respectively such that

$$u_1 = a^S, \quad b = a^{ST}; \quad v = a^{T_1}, \quad b = a^{T_1 S_1}, \quad \textit{i.e.} \quad a^{ST} = a^{T_1 S_1}.$$

* P. Dubreil, *loc. cit.*, 72.

Hence $STS_1^{-1}T_1^{-1}$ is in C_a, *i.e.* $ST = CT_1S_1$. Keeping a fixed and allowing S, T to range over G_{R_1}, G_{R_2} respectively, we have $G_{R_1}G_{R_2} \leqslant C_a G_{R_2} G_{R_1}$ and, allowing a to range over $\mathfrak{G}$, $G_{R_1}G_{R_2} \leqslant (\ldots C_a G_{R_2} G_{R_1} \ldots)$.

COROLLARY 1. *If $\mathfrak{G}$ is regular and if R_1, R_2 are associable, then $C_a = 1$, i.e., $G_{R_1}G_{R_2} = G_{R_2}G_{R_1} = [G_{R_1}, G_{R_2}] = G_{[R_1, R_2]}$.*

COROLLARY 2. *If $G_{R_2}G_{R_1}$ is distributive over $(\ldots C_a \ldots)$ and if R_1, R_2 are associable, then $G_{R_1}G_{R_2} = G_{R_2}G_{R_1}$.*

THEOREM 16. *If R_1 and R_2 are associable and minimal, then $G_{[R_1, R_2]} \leqslant C_a G_{R_1} G_{R_2}$ and $G_{R_1}G_{R_2} \leqslant C_a G_{[R_2, R_1]}$; $G_{R_2}G_{R_1} \leqslant C_a G_{[R_1, R_2]}$.*

Since R_1, R_2 are associable if $a \equiv b \pmod{[R_1, R_2]}$, it follows that u exists such that $a \equiv u \pmod{R_1}$ and $u \equiv b \pmod{R_2}$.

Since R_1, R_2 are minimal, so also is $[R_1, R_2]$. Let M be any element of $G_{[R_1, R_2]}$, hence $a \equiv a^M \pmod{[R_1, R_2]}$. Therefore S, T exist in G_{R_1}, G_{R_2} respectively such that $a \equiv a^S \pmod{R_1}$, $a^S \equiv a^M \pmod{R_2}$, *i.e.* $a^M = a^{ST}$, *i.e.* $M = C.ST$, where C is in C_a. Allowing M to range over $G_{[R_1, R_2]}$, we have $G_{[R_1, R_2]} \leqslant C_a G_{R_1} G_{R_2}$.

Again, let S, T be any elements in G_{R_1}, G_{R_2}; then $a \equiv a^S \pmod{R_1}$ $a^S \equiv a^{ST} \pmod{R_2}$. Hence $a \equiv a^{ST} \pmod{[R_1, R_2]}$. Therefore M exists in $G_{[R_1, R_2]}$ such that $a^{ST} = a^M$, *i.e.* $ST = C.M$, where C is in C_a. Allowing S, T to range over G_{R_1}, G_{R_2} respectively, we have $G_{R_1}G_{R_2} \leqslant C_a G_{[R_1, R_2]}$.

COROLLARY 1. *If R_1, R_2 are associable and minimal, then*

$$G_{R_1}G_{R_2} \leqslant [G_{R_1}, G_{R_2}] \leqslant (\ldots C_a G_{R_1} G_{R_2} \ldots).$$

COROLLARY 2. *If R_1, R_2 are associable and minimal, then*

$$G_{[R_1, R_2]} \leqslant (C_a G_{R_1} G_{R_2},\ C_a G_{R_2} G_{R_1}).$$

COROLLARY 3. *This, together with either the ascending or descending chain condition, and with the condition that (R_i, R_j) is minimal, is a sufficient condition for the set of minimal R_i to be a Dedekind structure.**

Definition 8. R_1 and R_2 are said to be *semi-consecutive* if, whenever $a \equiv b \pmod{R_1}$ and $b \equiv c \pmod{R_2}$, then either $a \equiv c \pmod{R_1}$ or $a \equiv c \pmod{R_2}$.

THEOREM 17. *If R_1 and R_2 are minimal, then a necessary and sufficient condition† for them to be semi-consecutive is that every element of $G_{R_1}G_{R_2}$ is in either $C_a G_{R_1}$ or $C_a G_{R_2}$.*

---

* Morgan Ward, *Bull. American Math. Soc.*, 45 (1939), 448.

† P. et M.-L. Dubreil, *loc. cit.*, 78, Theorem V.

Let S, T be any elements in G_{R_1}, G_{R_2} respectively and let a be any element in $\mathcal{G}$.

Then $a \equiv a^S \pmod{R_1}$, $a^S \equiv a^{ST} \pmod{R_2}$. Hence, if R_1 and R_2 are semi-consecutive, then

either $a \equiv a^{ST} \pmod{R_1}$ or $a \equiv a^{ST} \pmod{R_2}$,

i.e. either $a = a^{STS_1}$ or $a = a^{STT_1}$,

i.e. either $STS_1 = C_a$ or $STT_1 = C_a$, where c_a is in C_a,

i.e. every element in $G_{R_1} G_{R_2}$ is in either $C_a G_{R_1}$ or $C_a G_{R_2}$.

Conversely, assuming the conditions, let $a \equiv b \pmod{R_1}$, $b \equiv c \pmod{R_2}$ then S, T exist in G_{R_1}, G_{R_2} respectively such that $b = a^S$, $c = b^T = a^{ST}$, but $ST = CS_1$ or $ST = CT_1$, where C is in C_a. Hence either

$$c = a^{CS_1} = a^{S_1} \equiv a \pmod{R_1} \quad \text{or} \quad c = a^{CT_1} = a^{T_1} \equiv a \pmod{R_2},$$

i.e. R_1, R_2 are semi-consecutive.

COROLLARY 1. *This is also a necessary and sufficient condition for ensuring that, if* $(R_1, R_2) \subset R \subset R$, *then* R *and* R_2 *are associable*.*

COROLLARY 2. *This is also a sufficient condition for the structure-quotients to satisfy the second law of isomorphism, viz.,*

$$[R_1, R_2]/R_1 \cong R_2/(R_1, R_2).$$

2. *Correspondence between the set of all equivalence relations and the set of inertial automorphisms.*

Let Σ_i denote the set of all isotopic equivalence relations in $\mathfrak{R}$ having G_i as inertial group. These sets are disjoint and there is a (1, 1) correspondence $\Sigma_i \rightleftarrows G_i$.

The classes Σ_i may be defined as a number system by postulating that (Σ_i, Σ_j) is the complete set of cross-cuts (R_i, R_j), where R_i is in Σ_i and R_j is in Σ_j. Since $G_{(R_i R_j)} = (G_{R_i}, G_{R_j})$ is the same for all possible choices R_i, R_j, it follows that every element in (Σ_i, Σ_j) is in the same class $\Sigma_{ij} \to (G_{R_i}, G_{R_j})$. We define (Σ_i, Σ_j) to be Σ_{ij} and we denote the resulting system by $\mathfrak{R}/J$, where J denotes the set of all inertial sub-groups. Hence $\mathfrak{R}/J \cong J$ with respect to cross-cut.

The set is therefore a groupoid; it contains a zero, viz. the class Σ_E corresponding to 1; it contains a unit, viz. the class Σ_0 corresponding to $G(\mathcal{G})$.

* *Loc. cit.*, 83.

Also any automorphism of $\mathfrak{R}$ with respect to cross-cut (mod P) induces an automorphism of J with respect to cross-cut. For let S be an automorphism of $\mathfrak{R}$, *i.e.* $R_1 \to R_1^S$, $R_2 \to R_2^S$; then $(R_1, R_2) \to (R_1^S, R_2^S)$, and let $G_R \to G_{R^S}$. Then $G_{(R_1, R_2)} = (G_{R_1}, G_{R_2})$ and $G_{(R_1^S, R_2^S)} = (G_{R_1^S}, G_{R_2^S})$; hence $(G_{R_1}, G_{R_2}) \to (G_{R_1^S}, G_{R_2^S})$.

This correspondence will not be discussed further.

PART II.

Theorems depending on the law of multiplication.

So far, the facts that $\mathcal{G}$ is a groupoid and that $G(\mathcal{G})$ is a group of automorphisms have not been used. In this part the theorems involve the multiplicative properties of $\mathcal{G}$.

3. *Non-transitive classes.*

When R is minimal certain classes are mapped on themselves by all automorphisms:

(*a*) *Classes of idempotents.* If $a^2 = a$, then $a^S a^S = (a^2)^S = a^S$, *i.e.* a^S is also an idempotent for all automorphisms S. Hence, R being minimal, a class which contains an idempotent consists exclusively of idempotents. Certain sub-sets of these classes are also mapped on themselves:

(i) *Classes of left-zeros*; if $za = z$ for all a, then $z^S a^S = z^S$ for all a^S, *i.e.* $z^S b = z^S$ for all b, *i.e.* z^S is also a left-zero.

(ii) *Classes of right-zeros* similarly.

The sub-classes (i) and (ii) are distinct since, if $\mathcal{G}$ contains a right and also a left-zero, these are equal and consequently unique.

(iii) *Classes of left-units*; if $ea = a$, then $e^S a^S = a^S$, *i.e.* $e^S b = b$ for all b. Hence e^S is also a left-unit.

(iv) *Classes of right-units* similarly.

A class of right-units may coincide with a class of left-zeros.

(*b*) *Classes of non-homogeneous elements*, *i.e.* elements which are not products of others. A homogeneous element can never be mapped on a non-homogeneous one; for, if a is homogeneous, $a = bc$, then $a^S = b^S . c^S$ is also homogeneous for all automorphisms S.

Such classes are distinct from all other classes.

(c) *Classes of the centrum.* Let c be in the centrum, and let a be any element in $\mathcal{Q}$; then $ac = ca$. Hence $a^S c^S = c^S a^S$, *i.e.* $bc^S = c^S b$ for all b, *i.e.* c^S is in the centrum.

In general the centrum is not a sub-groupoid of $\mathcal{Q}$, since some weak form of the associative law is necessary for that to be the case.

(d) *Classes of the nucleus.* Let $\mathfrak{B}$ be *any* sub-groupoid and let $\mathfrak{K}$ be the nucleus. Then $\mathfrak{B}\mathfrak{K} = \mathfrak{K}\mathfrak{B}$. If, therefore, b is any element of $\mathcal{Q}$ and k any element of $\mathfrak{K}$, then $bk = k\bar{b}$, $\bar{b}$ being some power of b since $\mathfrak{B}$ may be taken to be a cyclic sub-groupoid. Hence $b^S k^S = k^S \bar{b}^S = k^S \bar{b}^{\bar{S}}$. Hence k^S belongs to the nucleus.

In general, a weak form of the associative law is necessary for the nucleus to be a sub-groupoid of $\mathcal{Q}$.

(e) *Neutral elements, i.e.* elements invariant under all automorphisms. Each of these is the sole member of a class if R is minimal. The neutral elements form a sub-groupoid; for, if u_1, u_2 are in $\mathfrak{N}$, then, if $u_1 u_2 = a$, we have $u_1 u_2 = a^S = a$ for all S, *i.e.* a is also neutral.

Special instances of congruence.

In most applications the congruences are special, *e.g.* in groups; the classes (mod R) are of two kinds, viz. classes of self-conjugate elements or coset expansion classes.

Definition 9. The *all* element O of $\mathfrak{R}$ is that in which there is only one class, *viz.* $\mathcal{Q}$ itself.

Definition 10. The *unit* element E of $\mathfrak{R}$ is that in which every element is the sole member of its class.

Evidently $G_O = G(\mathcal{Q})$ and $G_E = 1$.

Definition 11. An equivalence relation $\bar{R}$ such that $[R, \bar{R}] = O$, $(R, \bar{R}) = E$ is termed a *complement* of R.

$\bar{R}$ always exists; for the equivalence relation in which one class contains exactly one element from each class (mod R), the remaining elements of $\mathcal{Q}$ each constituting a class, is evidently a complement of R.

$\bar{R}$ is not in general unique.

In general, G_R, $G_{\bar{R}}$ are not permutable; for, if in R b is the only element in its class and $a \not\equiv b$ (mod $\bar{R}$) then, if $a \equiv u$ (mod R), $u \equiv b$ (mod $\bar{R}$), *i.e.* v cannot be found such that $a \equiv v$ (mod $\bar{R}$), $v \equiv b$ (mod R).

Also, in general, no $\bar{R}$ is minimal if R is minimal; otherwise O is minimal, *i.e.* $\mathcal{G}$ is transitive.

Evidently $(G_R, G_{\bar{R}}) = G_{(R, \bar{R})} = G_E = 1$, and, if $\mathcal{G}$ is transitive or if $[G_R, G_{\bar{R}}]$ is distributive with respect to $(\ldots C_a \ldots)$, then $G(\mathcal{G}) = [G_R, G_{\bar{R}}]$.

Definition 12. If $R_1 \subset R$ and $[R_1, \bar{R}_1] = R$, $(R_1, \bar{R}_1) = E$, then $\bar{R}_1$ is termed a *complement of* R_1 *in* R.

$\bar{R}_1$ always exists. Thus let $\bar{R}_1$ be the equivalence relation in which the classes are formed by taking exactly one element from each class (mod R_1) which lies in the same class (mod R), the remaining elements being the sole elements in the remaining classes (mod $\bar{R}_1$), *e.g.*

R_1:	R:	$\bar{R}_1$:
$a\ d\ g\ l$;	$a\ d$;	$a\ d\ b\ c\ e\ f\ k$
$b\ e\ k$	$b\ e$	$g\ l$
$c\ f$	$c\ f$	
	$g\ l$	
	k	

Then $[R_1, \bar{R}_1] = R$ and $(R_1, \bar{R}_1) = E$ and $[G_{R_1}, G_{\bar{R}_1}] \leqslant G_R$; $(G_{R_1}, G_{\bar{R}_1}) = 1$.

In general, $\bar{R}_1$ is not unique. This fact throws some light on properties of the inertial automorphisms. Let $R = [R_1, R_2] = [R_1, R_3]$; $(R_1, R_2) = (R_1, R_3) = E$. Then

$$\Big[R_2, (R_1, [R_2, R_3])\Big] \subset ([R_2, R_3], [R_1, R_2]) \subset (R, [R_2, R_3]) \subset [R_2, R_3].$$

Hence $R_4 = (R_1, [R_2, R_3])$ exists and

$$(R_2, R_4) = \Big(R_2, (R_1, [R_2, R_3])\Big) = \Big((R_1, R_2), [R_2, R_3]\Big) = (E, [R_2, R_3]) = E,$$

i.e. an inertial sub-group G_{R_4} exists such that

$$[G_{R_2}, G_{R_4}] \leqslant G_{[R_2, R_3]}, \quad (G_{R_2}, G_{R_4}) = 1.$$

Similarly
$$[G_{R_3}, G_{R_4}] \leqslant G_{[R_2, R_3]}, \quad (G_{R_3}, G_{R_4}) = 1.$$

If R_1, R_2, R_3 satisfy the Dedekind condition*, then

$$[R_2, R_4] = [R_2, R_3] = [R_3, R_4] \quad \text{and} \quad (R_2, R_4) = (R_3, R_4) = E.$$

Further, if $[G_{R_i}, G_{R_j}] = G_{[R_i, R_j]}$, then G_{R_4} is an Abelian group belonging to the centre of $[G_{R_2}, G_{R_3}]$†, for then

$$[G_{R_2}, G_{R_4}] = [G_{R_2}, G_{R_3}] = [G_{R_3}, G_{R_4}] \quad \text{and} \quad (G_{R_2}, G_{R_4}) = (G_{R_3}, G_{R_4}) = 1.$$

---

* O. Ore, *loc. cit.*, 412.

† O. Ore, *Duke Math. Journal*, 3 (1937), 172.

Definition 13. A *fundamental set* (mod R) is a maximal set of elements no one of which is the image of any other in G_R.

The classes of elements (mod $\bar{R}$) are examples of fundamental sets. The cases of special interest are those in which a fundamental set is associated with an automorphism in G_R. If F is a fundamental set, then so is F^S and, unless S belongs to G_F, F^S differs from F. F^S and F may have common elements, *e.g.* unit and neutral elements. Hence if it is desired to establish a correspondence between the elements of G_R and the fundamental sets it is necessary to postulate a convention by which such elements are allocated to the sets. Well-known examples are the fundamental regions of a group. The subject will not be discussed here.

4. *Coset expansions.*

So far, the fact that G_R is a group of automorphisms has not been used (except in § 3). In the following sections this fact is essential to the argument.

Let $R_{\mathfrak{B}}$ denote the equivalence relation in which $\mathfrak{G} = \sum_i \mathfrak{B}a_i + \mathfrak{N}$, where $\mathfrak{B}$ is a sub-groupoid and $\mathfrak{N}$ the set of all non-homogeneous elements in $\mathfrak{G}$. This is termed a *coset expansion* if:—

(*a*) *Every element of* $\mathfrak{G}$ *is in some coset, i.e.* if g is any homogeneous element of $\mathfrak{G}$, then b and a can be found, b in $\mathfrak{B}$, such that $g = ba$.

(*b*) *The cosets are disjoint, i.e.* if $b_1 a_1 = b_2 a_2$, then $\mathfrak{B}a_1 = \mathfrak{B}a_2$.

(*c*) *Every element in a coset may be taken as leader, i.e.* $\mathfrak{B}(ba) = \mathfrak{B}a$ for all b in $\mathfrak{B}$.

The class $\mathfrak{N}$ has no leader.

This condition implies a weak form of the associative law, viz.:

$$b_1(b_2 a) = ba, \qquad \text{(I)}$$

where b is an element in the sub-groupoid $\{b_1, b_2\}$ generated by b_1, b_2.

A stronger form of this associative law is

$$b_1(b_2 a) = (\bar{b}_1 b_2)\, a, \qquad \text{(II)}$$

$\bar{b}_1$ being some power (non-associative) of b_1.

THEOREM 18. *Any automorphism* z *of* $G_z(\mathfrak{B})$, *the group of decomposition of* $\mathfrak{B}$, *maps cosets into cosets.*

Let b be any element in $\mathfrak{B}$; then $(ba)^z = b^z a^z$; but, since z is in $G_z(\mathfrak{B})$, b^z is in $\mathfrak{B}$. Let $b^z = b_1$; then $(ba)^z = b_1 a^z$, *i.e.* every element in $(\mathfrak{B}a)^z$ is in $\mathfrak{B}a^z$ and, since the correspondence z is (1, 1), $(\mathfrak{B}a)^z = \mathfrak{B}a^z$.

THEOREM 19. *$G_{R_{\mathfrak{B}}}$ is a normal divisor of $G_z(\mathfrak{B})$.*

Let S, z be automorphisms in $G_{R_{\mathfrak{B}}}$ and $G_z(\mathfrak{B})$ respectively. Then $(\mathfrak{B}a)^{zSz^{-1}} = \mathfrak{B}a^{zSz^{-1}}$; but, since S is in $G_{R_{\mathfrak{B}}}$, $a^{zS} = (a^z)^S$ is in the same coset as a^z and, since every element in a coset may be taken as leader, $\mathfrak{B}a^{zS} = \mathfrak{B}a^z$. Hence $\mathfrak{B}a^{zSz^{-1}} = \mathfrak{B}a$, *i.e.* zSz^{-1} leaves every coset invariant and must therefore belong to $G_{R_{\mathfrak{B}}}$, which is therefore a normal divisor of $G_z(\mathfrak{B})$.

A well-known example is the Hilbert group of inertia in an algebraic number field, in which case $G_z(\mathfrak{B})/G_{R_{\mathfrak{B}}}$ is cyclic.

In the general case G_R is not a normal divisor of $G_z(\mathfrak{B})$, for although a^{zS} is in the same class as a^z it does not follow that $a^{zSz^{-1}}$ is in the same class as $(a^z)^{z^{-1}} = a$, for z^{-1} may split the class. The significance of the theorem is that when R determines a coset expansion, such splitting does not occur.

The non-homogeneous part is mapped onto itself by all automorphisms.

THEOREM 20. *Let $G_z(\mathfrak{B}) = G_{R_{\mathfrak{B}}} + \Sigma G_{R_{\mathfrak{B}}} S$; then every element in the coset $G_{R_{\mathfrak{B}}} S$ maps the coset $\mathfrak{B}a$ into the coset $\mathfrak{B}a^S$ and $\mathfrak{B}$ into $\mathfrak{B}$.*

Let T be any automorphism in $G_{R_{\mathfrak{B}}}$ and a any element in $\mathfrak{g}$. Then $a^{TS} = a_2{}^S$, where a^T and a_2 are in the same coset. But a is in the same coset as a^T, hence a_2 and a are in the same coset, *i.e.*, if a_2 is homogeneous, then $a_2 = ba$. Hence $a_2{}^S = b^S a^S = b_2 a^S$. Therefore, since $\mathfrak{B}(b_2 a^S) = \mathfrak{B}a^S$, $\mathfrak{B}a^{TS} = \mathfrak{B}a^S$ for all T in $G_{R_{\mathfrak{B}}}$.

In general, it is possible for elements in different cosets $G_{R_{\mathfrak{B}}} S_1$, $G_{R_{\mathfrak{B}}} S_2$ to map different cosets into $\mathfrak{B}a$, for $S_1 S_2^{-1}$ may leave a invariant, *i.e.* $S_1 = C_a S_2$, where C_a is in $G_z(\mathfrak{B})$ but not in $G_{R_{\mathfrak{B}}}$.

In certain instances the quotient groupoid $G_z(\mathfrak{B})/G_{R_{\mathfrak{B}}}$ is simply isomorphic* with the group of automorphisms of $\mathfrak{g}/\mathfrak{B}$. Suppose that the *quotient groupoid* $G/\mathfrak{B}$ exists, *i.e.* $\mathfrak{B}a \,.\, \mathfrak{B}c = \mathfrak{B}ac$. Then every element in $G_{R_{\mathfrak{B}}S}$ maps $\mathfrak{B}a$ into $\mathfrak{B}a^S$; $\mathfrak{B}c$ into $\mathfrak{B}c^S$ and $\mathfrak{B} \,.\, ac$ into $\mathfrak{B}(ac)^S$. Hence $\mathfrak{B}a^S \mathfrak{B}c^S = \mathfrak{B}(ac)^S$. Therefore the correspondence $\mathfrak{B}a \to \mathfrak{B}a^S$ is an automorphism of $\mathfrak{g}/\mathfrak{B}$ if it is a (1, 1) correspondence. Also if

$$S' = \begin{pmatrix} \mathfrak{B}a \\ \mathfrak{B}a^S \end{pmatrix}, \quad T' = \begin{pmatrix} \mathfrak{B}a \\ \mathfrak{B}a^T \end{pmatrix},$$

then $S'T' = (ST)'$. Hence $G_z(\mathfrak{B})/G_{R_{\mathfrak{B}}}$ is simply isomorphic with a subgroup of the group of automorphisms of $\mathfrak{g}/\mathfrak{B}$.

* Saunders MacLane, *Annals of Math.*, 40 (1939), 437.

Let T denote an automorphism on $\mathcal{G}/\mathfrak{B}$ in which $\mathfrak{B}a \to \mathfrak{B}a_T$, $\mathfrak{B}c \to \mathfrak{B}c_T$ and $\mathfrak{B} \to \mathfrak{B}$. Then $\mathfrak{B}a \,.\, \mathfrak{B}c = \mathfrak{B}ac$ and $\mathfrak{B}a_T\,\mathfrak{B}c_T = \mathfrak{B}a_T c_T$. Hence

$$\mathfrak{B}(ac)_T = \mathfrak{B}a_T c_T,$$

i.e. b_1, b_2 exist in $\mathfrak{B}$ such that $(ac)_T = b_1(a_T c_T)$, $a_T c_T = b(ac)_T$, *i.e.* $(ac)_T = b_1\big(b_2(ac)_T\big)$. In certain systems, *e.g.* in a ring of algebraic integers, we can deduce from this that $b_1 = b_2 =$ a left-hand unit. If so, then $(ac)_T = a_T c_T$; in this case the correspondence $a \to a_T$ is an automorphism T' on the leaders of the cosets of $\mathcal{G}/\mathfrak{B}$; and, if T' can be *extended* to an automorphism of $\mathcal{G}$, then $G_z(\mathfrak{B})/G_{R_{\mathfrak{B}}} \cong$ the group of automorphisms of $\mathcal{G}/\mathfrak{B}$.

THEOREM 21. $R_{(\mathfrak{B},\,\mathfrak{C})} \subset (R_{\mathfrak{B}},\, R_{\mathfrak{C}})$.

Let $\mathcal{G} = \mathfrak{B} + \mathfrak{B}g_2 + \ldots$, $\mathcal{G} = \mathfrak{C} + \mathfrak{C}h_2 + \ldots$ be coset expansions of $\mathcal{G}$ and let $\mathfrak{D} = (\mathfrak{B},\, \mathfrak{C})$. Then $\mathfrak{D}$ is a class of elements of $\mathcal{G}$ congruent mod $(R_{\mathfrak{B}},\, R_{\mathfrak{C}})$, for it consists of elements which are congruent mod $R_{\mathfrak{B}}$ and also mod $R_{\mathfrak{C}}$. Suppose that the coset expansion $\mathcal{G} = \mathfrak{D} + \mathfrak{D}l_2 + \ldots$ exists; then $\mathfrak{D}l_2$ consists of all elements of the form dl_2, where d is in both $\mathfrak{B}$ and $\mathfrak{C}$. Now l_2 is in some coset $\mathfrak{B}g_2$, *i.e.* $l_2 = bg_2$; hence $dl_2 = d(bg_2)$, and, from (c), $dl_2 = b_2 g_2$, where b_2 is in $\mathfrak{B}$ and g_2 depends only on l_2. Hence every element in $\mathfrak{B}l_2$ is in $\mathfrak{B}g_2$. Similarly, $\mathfrak{D}l_2 \leqslant \mathfrak{C}h_2$ and it is therefore in some complete class mod $(R_{\mathfrak{B}},\, R_{\mathfrak{C}})$. Hence $R_{(\mathfrak{B},\,\mathfrak{C})} \subset (R_{\mathfrak{B}},\, R_{\mathfrak{C}})$.

For equality a further postulate is necessary.

POSTULATE. *Weak form of the cancellation law.*

If $ba = ca$, then b and c both belong to $(\mathfrak{B},\, \mathfrak{C})$. A consequence of this is that $(a\mathfrak{B},\, a\mathfrak{C}) = a(\mathfrak{B},\, \mathfrak{C})$.

THEOREM 22. *If this cancellation law holds, then* $(R_{\mathfrak{B}},\, R_{\mathfrak{C}}) = R_{(\mathfrak{B},\,\mathfrak{C})}$.

By definition $a \equiv b\ [\mathrm{mod}\ (R_{\mathfrak{B}},\, R_{\mathfrak{C}})]$, if $a \equiv b\ (\mathrm{mod}\ R_{\mathfrak{B}})$ and also $(\mathrm{mod}\ R_{\mathfrak{C}})$.

Let $x = bg_2 = ch_2$ and let $b_1 g_2 = c_1 h_2 = y$. We have to prove that $x \equiv y\ (\mathrm{mod}\ R_{(\mathfrak{B},\,\mathfrak{C})})$.

x is in some coset $\mathfrak{D}l_2$; let $x = bg_2 = ch_2 = dl_2$; then, since $\mathfrak{D} \leqslant \mathfrak{B}$ and the cosets are disjoint, $\mathfrak{B}g_2 = \mathfrak{B}l_2$ and $\mathfrak{C}h_2 = \mathfrak{C}l_2$. Hence b_3 and c_3 exist in $\mathfrak{B}$ and $\mathfrak{C}$ respectively such that $b_1 g_2 = b_3 l_2$, $c_1 h_2 = c_3 l_2$, *i.e.* $b_3 l_2 = c_3 l_2$, and, by the cancellation law, b_3 and c_3 are both in $(\mathfrak{B},\, \mathfrak{C})$, *i.e.* in $\mathfrak{D}$. Hence $b_1 g_2 = c_1 h_2 = y = d_2 l_2$. Therefore $x \equiv y\ (\mathrm{mod}\ R_{\mathfrak{D}})$, *i.e.*

$$(R_{\mathfrak{B}},\, R_{\mathfrak{C}}) \subset R_{(\mathfrak{B},\,\mathfrak{C})}.$$

COROLLARY 1. *Under the conditions stated,* $(R_{\mathfrak{B}},\, R_{\mathfrak{C}}) = R_{(\mathfrak{B},\,\mathfrak{C})}$.

COROLLARY 2. $(G_{R_{\mathfrak{B}}},\, G_{R_{\mathfrak{C}}}) = G_{(R_{\mathfrak{B}},\, R_{\mathfrak{C}})} = G_{R_{(\mathfrak{B},\,\mathfrak{C})}}$.

THEOREM 23. *If the associative laws* $(ab)c = \bar{a}(bc)$; $a(bc) = (\bar{a}b)c$ *hold,* $\bar{a}$ *being some power of* a, *possibly depending on* b *and* c, *then the elements of order* n *of* $\mathfrak{M} = [\mathfrak{B}, \mathfrak{C}]$ *are of one of four forms, viz.* b; c; b *(a product of order* $n-1$ *at most);* c *(a product of order* $n-1$ *at most), where the order of a product is the number of elements in it.*

The elements of order 3 are $b_1(b_2c)$, $(b_1b_2)c = b_3c, b_1(cb_2)$, $(b_1c)b_2 = \bar{b}_1(cb_2)$, $c_1(c_2b)$, $(c_1c_2)b = c_3b$, $c_1(bc_2)$, $(c_1b)c_2 = \bar{c}_1(bc_2)$; $b_1(b_2b_3) = b_4$, $c_1(c_2c_3) = c_4$, $(b_1b_2)b_3 = b_5$, $(c_1c_2)c_4 = c_5$. Hence the theorem is true for products of orders 3 and 2. Assume that it is true for all products of orders less than n and consider any product m_n of order n. Since multiplication is a binary operation, $m_n = m_r m_s$, where $n = r+s$. By hypothesis m_r is of one of the forms b, c, bm_{r-1}, cm_{r-1} where m_{r-1} is at most of order $r-1$. Hence m_n is of one of the forms

$$bm_s, \quad cm_s, \quad (bm_{r-1})m_s = \bar{b}(m_{r-1}m_s) = \bar{b}(m_{r+s-1}) = \bar{b}m_{n-1} \quad \text{or} \quad \bar{c}m_{n-1},$$

i.e. the theorem is also true for all products of order n, and so generally.

The set of all such products $\mathfrak{M}$ constitutes the union $\mathfrak{M} = [\mathfrak{B}, \mathfrak{C}]$ for $\mathfrak{B} \leqslant \mathfrak{M}$, $\mathfrak{C} \leqslant \mathfrak{M}$. Let $\mathfrak{P}$ be any groupoid such that $\mathfrak{B} < \mathfrak{P}$, $\mathfrak{C} < \mathfrak{P}$; then evidently $\mathfrak{M} \leqslant \mathfrak{P}$. Further, $\mathfrak{M}$ is a sub-groupoid of $\mathfrak{G}$; for, if m_1, m_2 belong to $\mathfrak{M}$, so does m_1m_2. Hence $\mathfrak{M}$ is the union of $\mathfrak{B}$ and $\mathfrak{C}$.

THEOREM 24. *If the above-mentioned associative laws hold, then* $R_{[\mathfrak{B}, \mathfrak{C}]} \subset [R_{\mathfrak{B}}, R_{\mathfrak{C}}]$.

Let $x \equiv y \pmod{R_{[\mathfrak{B}, \mathfrak{C}]}}$; then m exists in $\mathfrak{M} = [\mathfrak{B}, \mathfrak{C}]$ such that $x = my$, where m must have one of the forms

$$b_1[c_1\{b_2(c_2 \dots b_rc_r)\}]; \quad b_1[c_1\{b_2(c_2 \dots b_r)\}], \quad c_1[b_1\{c_2(b_2 \dots c_rb_r)\}],$$

$$c_1[b_1\{c_2(b_2 \dots c_r)\}].$$

Let $u = b_1[c_1\{b_2(c_2 \dots)\}]$; then $x = [b_1[c_1\{b_2(c_2 \dots)\}]y$, *i.e.*

$$x \equiv [c_1\{b_2(c_2 \dots)\}]y \pmod{R_{\mathfrak{B}}},$$

$$[c_1\{b_2(c_2 \dots)\}]y \equiv b_2(c_2 \dots)y \pmod{R_{\mathfrak{C}}}, \quad \dots, \quad c_ry = y \pmod{R_{\mathfrak{C}}}.$$

Hence $x = y \pmod{[R_{\mathfrak{B}}, R_{\mathfrak{C}}]}$. Similarly for the other forms of m. Therefore if $x \equiv y \pmod{R_{[\mathfrak{B}, \mathfrak{C}]}}$, then $x \equiv y \pmod{[R_{\mathfrak{B}}, R_{\mathfrak{C}}]}$, *i.e.*

$$R_{[\mathfrak{B}, \mathfrak{C}]} \subset [R_{\mathfrak{B}}, R_{\mathfrak{C}}].$$

THEOREM 25. *If the above associative laws hold and* $\mathfrak{G}$ *is homogeneous, then* $[R_{\mathfrak{B}}, R_{\mathfrak{C}}] = R_{[\mathfrak{B}, \mathfrak{C}]}$.

Let $x \equiv y \bmod [R_{\mathfrak{B}}, R_{\mathfrak{C}}]$; then $a_1, a_2, \ldots, a_{2r-1}$ exist such that

$$x \equiv a_1 \pmod{R_{\mathfrak{B}}}, \quad a_1 \equiv a_2 \pmod{R_{\mathfrak{C}}}, \quad \ldots, \quad a_{2r-1} \equiv y \pmod{R_{\mathfrak{C}}}.$$

Since the associative law holds, every element of a coset may be taken as leader. Hence b_1 exists in $\mathfrak{B}$ such that $x = b_1 a_1$; c_1 exists in $\mathfrak{C}$ such that $a_1 = c_1 a_2$. Therefore $x = b_1 a_1 = b_1(c_1 a_2) = (b_1 c_1) a_2 = m_1 a_2$, where m_1 is in $[\mathfrak{B}, \mathfrak{C}]$. Proceeding thus, we have $x = my$, *i.e.* $x \equiv y \pmod{R_{[\mathfrak{B}, \mathfrak{C}]}}$. Hence $[R_{\mathfrak{B}}, R_{\mathfrak{C}}] \subset R_{[\mathfrak{B}, \mathfrak{C}]}$. Therefore $R_{[\mathfrak{B}, \mathfrak{C}]} = [R_{\mathfrak{B}}, R_{\mathfrak{C}}]$.

COROLLARY 1. $[G_{R_{\mathfrak{B}}}, G_{R_{\mathfrak{C}}}] \leqslant G_{[R_{\mathfrak{B}}, R_{\mathfrak{C}}]} \leqslant G_{R_{[\mathfrak{B}, \mathfrak{C}]}}$.

COROLLARY 2. *Since the non-homogeneous part of $\mathfrak{Q}$ is mapped on itself the theorem is true even if $\mathfrak{Q}$ is non-homogeneous.*

THEOREM 26. *If $\mathfrak{C} < \mathfrak{B}$ then $R_{\mathfrak{C}} \subset R_{\mathfrak{B}}$; and, if $\mathfrak{C}$ has a right-unit element of $\mathfrak{Q}$ and if $R_{\mathfrak{C}} \subset R_{\mathfrak{B}}$, then $\mathfrak{C} < \mathfrak{B}$.*

If $\mathfrak{C} < \mathfrak{B}$, then $\mathfrak{C}a < \mathfrak{B}a$, *i.e.* $R_{\mathfrak{C}} \subset R_{\mathfrak{B}}$.

If $R_{\mathfrak{C}} \subset R_{\mathfrak{B}}$, then $\mathfrak{C}$ is in some coset $\mathfrak{B}a$; hence, if c is any element in $\mathfrak{C}$, then b exists in $\mathfrak{B}$ such that $c = ba$. But every element in a coset can be taken as leader, and so $\mathfrak{B}a = \mathfrak{B}c$. Therefore $\mathfrak{C} < \mathfrak{B}c$ for all c. If, therefore, $\mathfrak{C}$ contains a right-unit of $\mathfrak{Q}$, $\mathfrak{C} < \mathfrak{B}$.

COROLLARY. *If $R_{\mathfrak{B}} = R_{\mathfrak{C}}$, then $\mathfrak{B} = \mathfrak{C}$.*

THEOREM 27. *If a set of sub-groupoids $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, ... is Dedekindian, then so are the corresponding equivalence relations $R_{\mathfrak{A}}$, $R_{\mathfrak{B}}$, $R_{\mathfrak{C}}$, ..., and conversely.*

With the above-mentioned conditions, let $\mathfrak{C} < \mathfrak{A}$; then $R_{\mathfrak{C}} \subset R_{\mathfrak{A}}$ and $(R_{\mathfrak{A}}, R_{[\mathfrak{B}, \mathfrak{C}]}) = (R_{\mathfrak{A}}, [R_{\mathfrak{B}}, R_{\mathfrak{C}}]) = R_{(\mathfrak{A}, [\mathfrak{B}, \mathfrak{C}])}$. Therefore, if the sub-groupoids belong to a Dedekind structure, $\mathfrak{C} = (\mathfrak{A}, [\mathfrak{B}, \mathfrak{C}])$, *i.e.*

$$R_{\mathfrak{C}} = (R_{\mathfrak{A}}, [R_{\mathfrak{B}}, R_{\mathfrak{C}}]),$$

i.e. the equivalence relations are also Dedekindian.

Conversely, let $R_{\mathfrak{A}} \subset R_{\mathfrak{C}}$, then $R_{\mathfrak{C}} = (R_{\mathfrak{A}}, [R_{\mathfrak{B}}, R_{\mathfrak{C}}])$ if the equivalence relations are Dedekindian, *i.e.* $R_{\mathfrak{C}} = R_{(\mathfrak{A}, [\mathfrak{B}, \mathfrak{C}])}$, therefore $\mathfrak{C} = (\mathfrak{A}, [\mathfrak{B}, \mathfrak{C}])$, *i.e.* the sub-groupoids are also Dedekindian.

5. *Correspondences between the sub-groups of $G(\mathfrak{Q})$ and the sub-groupoids of $\mathfrak{Q}$.*

The set of all automorphisms which leave every element of a set Σ or sub-groupoid $\mathfrak{B}$ invariant is a sub-group of $G(\mathfrak{Q})$, viz. G_Σ or $G_{\mathfrak{B}}$.

THEOREM 28. *The set of all elements of $\mathcal{G}$, each of which is invariant under an automorphism S, either is void or is a sub-groupoid of $\mathcal{G}$.*

Let $a = a^S$, $b = b^S$; then $ab = a^S b^S = (ab)^S$, *i.e.* ab also belongs to the set, which is therefore a sub-groupoid of $\mathcal{G}$.

COROLLARY. *The set of all elements of $\mathcal{G}$, each of which is invariant under every automorphism of a given set Σ or a given sub-group H of $G(\mathcal{G})$, either is void or is a sub-groupoid of $\mathcal{G}$.*

Denote this sub-groupoid by $\mathcal{G}_H$. If $\mathcal{G}$ contains a modulus e, then, since $e^S = e$ for all automorphisms, $\mathcal{G}_H$ is non-void. Hence, if we regard the void set as a sub-groupoid, there exist correspondences

$$\mathfrak{B} \to G_{\mathfrak{B}}, \tag{1}$$

$$H \to \mathcal{G}_H. \tag{2}$$

Evidently, if $\mathfrak{B} \to G_{\mathfrak{B}}$, $G_{\mathfrak{B}} \to \mathcal{G}_{G_{\mathfrak{B}}}$, it does not follow in general that $\mathfrak{B} = \mathcal{G}_{G_{\mathfrak{B}}}$. All that can be asserted is that $\mathfrak{B} \leqslant \mathcal{G}_{G_{\mathfrak{B}}}$. Similarly, if $H \to \mathcal{G}_H$, $\mathcal{G}_H \to G_{\mathcal{G}_H}$, then $H \leqslant G_{\mathcal{G}_H}$.

The most interesting cases are those in which a Galois theory exists, viz., those in which $\mathfrak{B} = \mathcal{G}_{G_{\mathfrak{B}}}$ and $H = G_{\mathcal{G}_H}$.

THEOREM 29. *If $\mathfrak{B} < \mathfrak{C}$, then $G_{\mathfrak{C}} \leqslant G_{\mathfrak{B}}$.*

For every automorphism which leaves each element of $\mathfrak{C}$ unaltered also leaves each element of $\mathfrak{B}$ unaltered.

THEOREM 30. *If $H < K$, then $\mathcal{G}_K \leqslant \mathcal{G}_H$.*

For every element invariant under every automorphism of K is also invariant under every automorphism of H.

THEOREM 31. *If $\mathfrak{B} \simeq \mathfrak{C} = \mathfrak{B}^S$, then $G_{\mathfrak{B}} = S G_{\mathfrak{C}} S^{-1}$.*

Let $c = b^S$, $b = c^{S^{-1}}$ and let T be any automorphism in $G_{\mathfrak{C}}$. Then $b = c^{S^{-1}} = c^{TS^{-1}} = b^{STS^{-1}}$. Hence, for all b in $\mathfrak{B}$, $b = b^{STS^{-1}}$, *i.e.* STS^{-1} is in $G_{\mathfrak{B}}$ and, if we allow T to range over $G_{\mathfrak{C}}$, $SG_{\mathfrak{C}} S^{-1} \leqslant G_{\mathfrak{B}}$. Similarly $S^{-1} G_{\mathfrak{C}} S \leqslant G_{\mathfrak{B}}$. Therefore $G_{\mathfrak{B}} = S G_{\mathfrak{C}} S^{-1}$.

COROLLARY. *If $\mathfrak{B} \simeq \mathfrak{C}$ are sub-groupoids of $\mathcal{G}$, it does not follow that all the correspondences which establish the isomorphism are automorphisms of $\mathcal{G}$. We can, however, assert that $G_{\mathfrak{B}} \simeq G_{\mathfrak{C}}$.*

THEOREM 32. *If H, K are conjugate sub-groups of $G(\mathcal{G})$, then $\mathcal{G}_H \simeq \mathcal{G}_K$.*

Let P be any automorphism in H; then, given S, Q exists in K such that $SP = QS$. Hence, if a is any element in $\mathcal{G}_K$,

$$(a^S)^P = a^{SP} = a^{QS} = (a^Q)^S = a^S,$$

i.e. a^S is invariant under all the automorphisms in H and therefore belongs to $\mathcal{G}_H$. Hence if a is in $\mathcal{G}_K$, a^S is in $\mathcal{G}_H$ and, since S is an automorphism, the correspondence $a \to a^S$ between the elements of $\mathcal{G}_K$ and $\mathcal{G}_H$ is (1, 1). It is also an isomorphism for $(ab) \to (ab)^S = a^S . b^S$, *i.e.* the correspondence is isomorphic and $\mathcal{G}_H \cong \mathcal{G}_K$, *i.e.* $\mathcal{G}_H = \mathcal{G}_K{}^S$ and $\mathcal{G}_K = \mathcal{G}_H^{S^{-1}}$.

COROLLARY 1. $\mathcal{G}_{G_{\mathfrak{B}}} = \mathcal{G}^{S^{-1}}_{G_{\mathfrak{B}^S}}$, $\mathcal{G}^{S}_{G_{\mathfrak{B}}} = \mathcal{G}_{G_{\mathfrak{B}^S}}$.

COROLLARY 2. *If* $\mathfrak{B} = \mathcal{G}_{G_{\mathfrak{B}}}$, *then* $\mathfrak{B}^S = \mathcal{G}_{G_{\mathfrak{B}^S}}$.

6. *Correspondence between the sub-groups of* $G(\mathcal{G})$ *and the inertial groups.*

An example shows that not every sub-group of $G(\mathcal{G})$ and not even every normal sub-group is an inertial group:

$\mathcal{G} =$

	a	b	c
a	a	c	b
b	c	b	a
c	b	a	c

$G(\mathcal{G}) = 1$, (ab), (bc), (ca), (abc), (acb), but 1, (abc), (acb) is not inertial.

A correspondence can, however, be established. Let A be any sub-group of $G(\mathcal{G})$ and let R_A denote the equivalence relation in which $a \equiv b$ if $b = a^S$, where S is in A, *i.e.* each class (mod R_A) is a set of elements conjugate under A. Corresponding to A there is the inertial group G_{R_A}, *i.e.* to every sub-group A there corresponds one, and only one, inertial group G_{R_A}, and every inertial group is the correspondent of at least one sub-group of $G(A)$.

THEOREM 33. *If* $A < B$, *then* $G_{R_A} \leqslant G_{R_B}$.

Since $A < B$, if $a \equiv b \pmod{R_A}$, it follows that $a \equiv b \pmod{R_B}$; hence $R_A \subset R_B$, and therefore $G_{R_A} \leqslant G_{R_B}$.

THEOREM 34. $G_{R_{(A,\,B)}} \leqslant G_{(R_A,\,R_B)}$.

For $R_{(A,\,B)} \subset R_A$, $R_{(A,\,B)} \subset R_B$; hence $R_{(A,\,B)} \subset (R_A, R_B)$. Therefore

$$G_{R_{(A,\,B)}} \leqslant G_{(R_A,\,R_B)} \leqslant (G_{R_A}, G_{R_B}).$$

THEOREM 35. $[G_{R_A}, G_{R_B}] \leqslant G_{R_{[A,\,B]}}$; $[R_A, R_B] = R_{[A,\,B]}$.

Let $a \equiv b \pmod{[R_A, R_B]}$; hence $u_1, u_2 \ldots, u_{2s-1}$ exist such that $a \equiv u_1 \pmod{R_A}$, $u_1 \equiv u_2 \pmod{R_B}$, $\ldots$, $u_{2s-1} \equiv b \pmod{R_B}$. Hence

$$S_1, S_2 \ldots, S_n; \; T_1, T_2, \ldots, T_n$$

exist in A and B respectively such that

$$a \equiv a^{S_1} \pmod{R_A}, \quad a^{S_1} \equiv a^{S_1 T_1} \pmod{R_B}, \quad \ldots, \quad a^{S_1 T_1 S_2 T_2 \ldots S_s} \equiv b \pmod{R_B},$$

i.e. $b = a^{S_1 T_1 S_2 T_2 \ldots S_s T_s} = a^M$, where M is in $M = [A, B]$. Hence

$$[R_A, R_B] \subset R_{[A, B]}.$$

Next, let $a \equiv b \pmod{R_{[A, B]}}$; then $M = S_1 T_1 \ldots S_s T_s$ exists such that $b = a^M = a^{S_1 T_1 \ldots S_s} \pmod{R_B}$, $a^{S_1 T_1 \ldots S_s} \equiv a^{S_1 T_1 \ldots T_{s-1}} \pmod{R_A}$, and so on. Hence $a \equiv b \pmod{[R_A, R_B]}$, *i.e.* $[R_A, R_B] \subset R_{[A, B]}$. Therefore

$$[R_A, R_B] = R_{[A, B]}.$$

Hence $[G_{R_A} G_{R_B} \leqslant G_{[R, R]}] \leqslant G_{R[A, B]}$.

COROLLARY. $G_{R_{(A, B)}} \leqslant (G_{R_A}, G_{R_B}) \leqslant [G_{R_A}, G_{R_B}] \leqslant G_{R_{[A, B]}}$.

7. *Algebraic automorphisms.*

In many instances a correspondence between the elements of $\mathcal{G}$ can be established algebraically. Since we are concerned only with automorphisms, such correspondences must be (1, 1); we therefore discuss only those algebraic types which are linear in spite of the fact that in many instances equations of higher order may still yield (1, 1) correspondences.

Type I. $zu_S = u_S z^S$ with its dual $u_S z = z^S u_S$.

These generalise the inner automorphisms of associative algebra.

Type II. $zv_S = z^S$ with its dual $v_S z = z^S$.

These lead in general only to the trivial unit automorphism in associative algebra.

(*a*) *Type* I.

THEOREM 36. *The set of all numbers z which are such that, for all elements a, b in $\mathcal{G}$, $(ab)z = a(bz)$ form an associative characteristic sub-groupoid $\mathcal{G}_*$ of $\mathcal{G}$.*

Let x, z be two such elements; then

$$(ab)(xz) = [(ab)x]z = [a(bx)]z = a[(bx)z] = a[b(xz)].$$

Hence xz also belongs to the set, which is therefore a sub-groupoid $\mathcal{G}_v$ of $\mathcal{G}$. Further, it is associative; for, if x, y, z belong to it, then $(xy)z = x(yz)$.

COROLLARY. *Similarly, the numbers z for which $(az)b = a(zb)$ and the numbers z for which $z(ab) = (za).b$ form associative sub-groupoids $\mathcal{G}_c$, $\mathcal{G}_l$ respectively.*

Definition 14. The cross-cut $(\mathcal{G}_v, \mathcal{G}_l, \mathcal{G}_c)$ is denoted by U and termed the *associative-nucleus of* $\mathcal{G}$.

It may, of course, be void; but, if not and if u belongs to it, so does u^S. Hence $U = U^S$ for all automorphisms S.

Consider now the correspondence $x \to x^S$ determined by the equation $xu_S = u_S x^S$, u_S being in U. Then

$$(xy)u_S = x(yu_S) = x(u_S y^S) = (xu_S)y^S = u_S(x^S)y^S = u_S(x^S y^S),$$

i.e. $xy \to x^S y^S$. Hence, if the correspondence is (1, 1), it is an automorphism of $\mathcal{G}$.

For the correspondence to be (1, 1) it is necessary that

(i) both the equations $zu_S = u_S x$, $xu_S = u_S z$ must be solvable in $\mathcal{G}$ for x.

(ii) if $u_S a = u_S b$ or if $au_S = bu_S$, then $a = b$.

THEOREM 37. *The elements of U satisfying conditions* (i) *and* (ii) *are groupoids.*

Let u_S, u_T satisfy the conditions; then, in a homogeneous groupoid, z exists such that $x(u_S u_T) = (u_S u_T)z$, *i.e.*

$$(xu_S)u_T = u_S(u_T z) = (u_S x^S)u_T = u_S(x^S u_T)$$

and, since the associative law holds, $x^S u_T = u_T z$; but x^S is uniquely determined by u_S and x, so that z is uniquely determined by u_S, u_T, x.

Definition 15. With the conditions specified the automorphisms induced by the algebraic equation

$$zu_S = u_S z^S \tag{1}$$

are termed automorphisms of type I.

THEOREM 38. *The set J of all automorphisms of type I is a groupoid consisting of elements of $G(\mathfrak{Q})$.*

$zu_S = u_S z^S$, hence $(zu_S)u_T = (u_S z^S)u_T$, *i.e.*

$$z(u_S u_T) = u_S(z^S u_T) = u_S(u_T z^{ST}) = (u_S u_T) z^{ST}.$$

Hence the correspondence $z \to z^{ST}$ is induced by the correspondence $z(u_S u_T) = (u_S u_T) z^{ST}$. Hence the automorphisms of type I form a groupoid.

COROLLARY 1. *If U contains elements of the centrum, these elements induce the identical automorphism E, and the groupoid J contains E.*

COROLLARY 2. *If J contains E, then $\mathfrak{Q}$ has a centrum.*

The same correspondence may possibly be established by more than one element u_S of U; for, if c is any element of the centrum of $\mathfrak{Q}$ (assumed non-void), then $z(u_S c) = (zu_S)c = (u_S z^S)c = u_S(z^S c) = u_S(cz^S) = (u_S c) z^S$, *i.e.* $u_S c$ induces the same automorphism as u_S.

Conversely, in a homogeneous groupoid, let $zu_S = u_S z^S$, $zx = xz^S$; then c can be found such that $x = u_S c$, and hence $z(u_S c) = (u_S c) z^S$, *i.e.*

$$(zu_S)c = u_S(cz^S) = (u_S z^S)c = u_S(z^S c);$$

and, if the cancellation law holds, $cz^S = z^S c$ for all z, *i.e.* c belongs to the centrum. Hence we have

THEOREM 39.
$$u_S u_T = h_{S,T} u_{ST}, \tag{2}$$

where $h_{S,T}$ belongs to the centrum of $\mathfrak{Q}$.

The identical automorphism of type I, $zu_S = u_S z$, requires u_S to be in the centrum of $\mathfrak{Q}$; and, conversely, every element in the centrum induces the unit automorphism. Let $\mathfrak{C}$ denote those elements of U which are in the centrum.

THEOREM 40. *$U/\mathfrak{C}$ exists and is simply isomorphic with the set J algebraic automorphisms of type I.*

If we assume that the coset expansion $U = \mathfrak{C} + \mathfrak{C}u_S + \ldots$ exists, then, since $\mathfrak{C}u_S \mathfrak{C}u_T = \mathfrak{C}u_S u_T = \mathfrak{C}u_{ST}$, $U/\mathfrak{C}$ exists as a quotient-groupoid. The correspondence $S \to \mathfrak{C}u_S$ is (1, 1) and evidently $ST \to \mathfrak{C}u_{ST}$. Hence

$$U/\mathfrak{C} \cong J.$$

This generalizes the theorems about inner automorphisms in the theory of groups and linear associative algebras.

Factor-sets.

If in the preceding section we restrict z to lie in a sub-groupoid $\mathfrak{B}$ instead of ranging over $\mathfrak{G}$, $h_{S,T}$ is no longer in the centrum of $\mathfrak{G}$ but is merely commutative with every element of $\mathfrak{B}$. Let us then use (2) and

$$u_S u_T = u_{ST} g_{S,T} \tag{3}$$

to define $g_{S,T}$, $h_{S,T}$; then

$$u_{ST} g_{S,T} = h_{S,T} u_{ST} = u_{ST} (h_{S,T})^{ST},$$

and, since the cancellation law is assumed,

$$g_{S,T} = (h_{S,T})^{ST}. \tag{4}$$

THEOREM 41. $$g_{S,TR} g_{T,R} = g_{ST,R} (g_{S,T})^R. \tag{5}$$

$$(u_S u_R) u_T = u_S (u_R u_T);$$

$$(u_{ST} g_{S,T}) u_R = u_{ST} (g_{S,T} u_R) = u_{ST} [u_R (g_{S,T})^R] = u_{ST} u_R (g_{S,T})^R$$

$$= u_{S,TR} g_{ST,R} (g_{S,T})^R.$$

Also

$$u_S (u_T u_R) = u_S (u_{TR} g_{T,R}) = (u_S u_{TR}) g_{T,R} = u_{S,TR} (g_{S,TR} g_{T,R}).$$

Hence $$u_{STR} g_{ST,R} (g_{S,T})^R = u_{STR} (g_{S,TR} g_{T,R}),$$

and, since the cancellation law holds,

$$g_{ST,R} (g_{S,T})^R = g_{S,TR} g_{T,R}.$$

Definition 16. The set $g = \{g_{S,T}\}$ of numbers satisfying (5) is termed a *factor-set.*

Other properties of factor-sets will not be discussed here since they can be still further generalized by replacing the associative law by a much weaker postulate involving automorphisms. It is noteworthy, however, that the corresponding theory* for linear associative algebras can be generalized.

(*b*) *Algebraic automorphisms of type* II.

The correspondence $z \to z^S$ determined by $z v_S = z^S$ is (1, 1) if, whenever $z v_S = x v_S$, $z = x$, *i.e.* if the left-hand cancellation law holds, and if the equation $z = x v_S$ always has a unique solution when v_S and z are fixed.

---

* M. Deuring, *Algebren.*

It is an automorphism if the associative law

$$(zx)\,v_S = (zv_S)(xv_S) \qquad \text{(I)}$$

holds.

Such automorphisms, except the identical automorphism, do not exist in associative algebra; for, in such algebras,

$$z(v_S\,v_T) = (zv_S)\,v_T \quad \text{implies that} \quad zv_S{}^T = z^{T^{-1}ST} = z^{ST} \quad \text{for all } z,$$

i.e. T leaves every element unaltered and is therefore 1, the unit automorphism.

THEOREM 42. *The algebraic automorphisms of type II form a groupoid.*

Let $zv_S = z^S$, $zv_T = z^T$ for all z. Then

$$(zv_S)\,v_T = z^T v_S{}^T = (zv_S)^T = z^S\,v_T = (z^S)^T = z^{ST}.$$

Hence $xv_S{}^T = x^{T^{-1}ST}$ for all x in $\mathcal{G}$. Therefore $zv_S{}^T = z^{T^{-1}ST} = z(v_S\,v_T)$, *i.e.* $v_S\,v_T$ determines the automorphism $z \to z^{T^{-1}ST}$; and, since it is unique,

$$v_S\,v_T = v_{T^{-1}ST}. \qquad \text{(II)}$$

Hence
$$v_R(v_S\,v_T) = v_R^{T^{-1}ST}; \quad (v_R\,v_S)\,v_T = v_R^{ST}. \qquad \text{(III)}$$

Next,
$$(zx)v_S\,v_T = (zx)\,v_S{}^T = (zx)^{T^{-1}ST}$$

and
$$\{z(v_S\,v_T)\}\{x(v_S\,v_T)\} = z^{T^{-1}ST}\,x^{T^{-1}ST} = (zx)^{T^{-1}ST}.$$

Hence
$$(zx)(v_S\,v_T) = \{z(v_S\,v_T)\}\{x(v_S\,v_T)\}.$$

Therefore $v_S\,v_T = v_{T^{-1}ST}$ belongs to the set, which is therefore a groupoid.

Evidently the identical automorphism $zv_E = z$ for all z requires v_E to be a right-unit, and any right-unit induces the identical automorphism.

Such right-units form a groupoid, for $z(v_E\,v_{E'}) = zv_E = z$; hence, if v_E, $v_{E'}$ are right-units, so also is $v_E\,v_{E'}$.

Let $\mathfrak{E}_r$ denote the right-unit sub-groupoid of $\mathcal{G}$ (assumed non-void).

THEOREM 42. *If $\mathfrak{V} = \{v_S\}$ is the groupoid which induces the groupoid K of algebraic automorphisms of type II, then $\mathfrak{V}/\mathfrak{E}_r \cong K$.*

Let $\mathfrak{V}/\mathfrak{E}_r = \mathfrak{E}_r + v\mathfrak{E}_r + \ldots$ exist as a coset expansion; then

$$v_1\mathfrak{E}_r \,.\, v_2\mathfrak{E}_r = v_1 v_2 \mathfrak{E}_r,$$

hence $V/\mathfrak{E}_r \cong K$.

COROLLARY. *V is characteristic, for $v_S{}^T = v_{T^{-1}ST} = v_S\,v_T$, i.e. $v_S{}^T$ is in V for all v_S and for all automorphisms T of type II.*

An automorphism may be of both types I and II; for, if $zu_S = u_S z^S$, $zv_S = z^S$, then $zu_S = u_S(zv_S) = (u_S z)v_S = u_S{}^S z^S = u_S z^S$, *i.e.*, since the cancellation law holds, $u_S = u_S{}^S$. Hence $u_S v_S = u_S{}^S = u_S$.

Conversely, if $u_S v_S = u_S$, then, v_S being of type II, $u_S{}^S = u_S$. Also $v_S^{S^{-1}} u_S = u_S v_S = u_S$, *i.e.* $v_S u_S{}^S = u_S{}^S$, *i.e.* $v_S u_S = u_S$; therefore

$$v_S{}^S u_S{}^S = u_S{}^S = v_S{}^S u_S = u_S,$$

i.e. $v_S{}^S = v_S$. Hence S is in G_U and in G_V and therefore in (G_u, G_v).

Pairs of elements may also induce automorphisms, *e.g.* $(u_S z)v_S = u_S(z^S v_S)$ induces an automorphism if $[u_S(xz)]v_S = u_S[(x^S z^S)v_S]$, and $(u_S z)v_S = z^S$ induces an automorphism if $[u_S(xz)]v_S = [(u_S x)v_S][(u_S z)v_S]$.

8. *Summary.*

In this paper contact has been made with the principal theorems relating to the automorphisms of associative linear algebra which are shown to be special cases of theorems in which the full associative law is not assumed. These theorems can be widely extended and specialized by postulating various weak forms of the associative laws.

University College,
Swansea.

THE CONTACT NET OF QUADRICS

By W. L. EDGE.

[Received 6 November, 1940.—Read 17 December, 1940.]

It is proposed to obtain, in the following pages, some properties of a net of quadric surfaces which is so specialised that its eight base points coincide in pairs; in other words, the quadrics of the net circumscribe a tetrahedron T and have a definite tangent line at each vertex of T. Such a net may perhaps fittingly be called a *contact net*, and it is so described in this paper.

Suppose then that A, B, C, D are the base points of a contact net $\mathfrak{C}$, with the fixed tangent lines α, β, γ, δ respectively. Four such lines are always tangents of a pencil, but not generally of a net, of quadrics. The figure cannot be set up arbitrarily, and the restriction to which it is subjected is expressed by the following

THEOREM. *A necessary and sufficient condition for the lines* α, β, γ, δ *to be tangent lines, at the vertices* A, B, C, D *of a tetrahedron* T, *of a net of quadrics, is that their respective intersections* A', B', C', D' *with the opposite faces of* T *form, with the vertices of* T, *a pair of Möbius tetrads.*

Thus it appears that the contact net is closely related to the Möbius net, and that until some account has been given of its properties the geometry of the Möbius configuration is not complete.

Now there is also an intimate relationship, between the contact net and the Möbius net, apart from that expressed by the theorem; for each net is derivable from the other by means of a simple Cremona transformation. Indeed, if one of the tetrads A, B, C, D of a Möbius configuration is taken to be the set of vertices of the tetrahedron of reference T for a set of homogeneous coordinates x, y, z, t, it is only necessary to replace these coordinates by their reciprocals; whereupon the tetrad A', B', C', D', whose points lie one in each of the faces of T, is made to coincide with the other, the transformation of A' being associated with a definite direction a through A, and so on. And the quadrics of the

Möbius net, passing through the eight points $A, B, C, D, A', B', C', D'$, are transformed into quadrics which circumscribe T and have $\alpha, \beta, \gamma, \delta$ as tangent lines at its respective vertices.

The properties of the Möbius net have been the subject of a preceding communication*, which will be referred to here as M. T.; so that material from which properties of $\mathcal{C}$ are immediately deducible by means of this transformation, which will be denoted by $\mathcal{S}$, lies ready to hand. Thus it is found, in § 2, that the cones of $\mathcal{C}$ fall into two families (a fact which has been alluded to† before), and that the envelope of the cones of either of the two families is a quartic scroll. These two quartic scrolls are a cardinal feature of the configuration, which may be built up from either of them; this is the standpoint of §§ 4–6, where it is shown how the contact net plays an essential part in the geometry of the quartic scroll, and where, in § 5, it is emphasised that, through its agency, quartic scrolls occur in pairs.

Yet, although properties of the contact net are so readily obtainable from those of the Möbius net once these latter are on record, it should not be forgotten that the contact net exists in its own right. So, before introducing the Möbius net, we give a proof of the theorem from first principles which displays its algebraical content to advantage. This is not pursued further; but, the theorem having been established independently, the canonical form I can be obtained as a corollary and properties of the net derived without the intervention of the Cremona transformation.

1. Take T as tetrahedron of reference for a system of homogeneous coordinates x, y, z, t; then the equations of $\alpha, \beta, \gamma, \delta$ have the respective forms

$$\frac{y}{m_1}=\frac{z}{n_1}=\frac{t}{p_1}; \quad \frac{z}{n_2}=\frac{t}{p_2}=\frac{x}{l_2}; \quad \frac{t}{p_3}=\frac{x}{l_3}=\frac{y}{m_3}; \quad \frac{x}{l_4}=\frac{y}{m_4}=\frac{z}{n_4}.$$

The matrix

$$M=\begin{pmatrix} . & m_1 & n_1 & p_1 \\ l_2 & . & n_2 & p_2 \\ l_3 & m_3 & . & p_3 \\ l_4 & m_4 & n_4 & . \end{pmatrix}$$

which is formed from the direction ratios of these four lines has, for the

* "The net of quadric surfaces associated with a pair of Möbius tetrads", *Proc. London Math. Soc.* (2), 41 (1936), 337–360.

† *Acta Mathematica*, 64 (1934); see pp. 237 and 238.

constituents of any one row, the coordinates of one of the four points A', B', C', D'. The vanishing of the four principal elements in M expresses the fact that these four points lie one in each of the faces of T. If they form a pair of Möbius tetrads with the vertices of T, the plane of any three of them passes through a vertex of T; the conditions for this to happen are that the four principal minors of M should vanish. Thus, in order to prove the theorem, the vanishing of these principal minors must be established.

Let the equations of three linearly independent quadrics which circumscribe T be

$$q_i \equiv f_i yz + g_i zx + h_i xy + u_i xt + v_i yt + w_i zt = 0 \quad (i = 1, 2, 3).$$

If every quadric $\xi q_1 + \eta q_2 + \zeta q_3 = 0$ touches α at A, then

$$m_1 h_i + n_1 g_i + p_1 u_i = 0 \quad \text{for} \quad i = 1, 2, 3;$$

hence the determinant (hgu) vanishes, its columns being linearly dependent according to the multipliers m_1, n_1, p_1. Corresponding considerations for the other vertices of T show that, when the quadrics of the net all have α, β, γ, δ for tangent lines at A, B, C, D, the relations

$$(hgu) = (fhv) = (gfw) = (uvw) = 0 \qquad (1.1)$$

must be satisfied. The three columns of any one of these determinants are linearly dependent according to multipliers which occur in one of the rows of M, and all these relations of linear dependence are given by the matrix equation

$$\begin{pmatrix} f_1 & g_1 & h_1 & u_1 & v_1 & w_1 \\ f_2 & g_2 & h_2 & u_2 & v_2 & w_2 \\ f_3 & g_3 & h_3 & u_3 & v_3 & w_3 \end{pmatrix} \begin{pmatrix} . & n_2 & m_3 & . \\ n_1 & . & l_3 & . \\ m_1 & l_2 & . & . \\ p_1 & . & . & l_4 \\ . & p_2 & . & m_4 \\ . & . & p_3 & n_4 \end{pmatrix} = 0.$$

Let this be written

$$FL = 0.$$

We now appeal to the known fact* that the sum of the ranks of F

* If X denotes a column vector of six elements, the equation $FX = 0$ has three linearly independent solutions. Thus a matrix such as L, formed by juxtaposition of solutions X, has rank 3 at most. Its rank is actually equal to 3 when its columns include a *complete* solution of $FX = 0$, and this the columns of L actually do. The distribution of zero elements in L makes it manifest that no three of its columns can be linearly dependent.

and L is 6 (the number of columns in F and of rows in L). Since the three quadratic forms q_i are presumed to be linearly independent, the rank of F is 3; wherefore the rank of L must also be 3. Hence every determinant of four rows and columns which occurs in L must vanish.

The vanishing of these determinants yields the desired conditions. For such a determinant is obtained when any two rows of L are omitted; in particular, two of the three rows containing the non-vanishing constituents of any one column may be omitted, and so it follows that four three-rowed determinants which, upon examination, are seen to be the four principal minors of M, must vanish.

It only remains to say, in order to complete the proof of the theorem, that not only do the principal minors of M vanish in consequence of L being of rank 3, but that, conversely, the rank of L reduces to 3 provided that the four principal minors of M vanish. For these conditions, by the manner in which they were found, ensure that twelve of the fifteen four-rowed determinants of L vanish; that the remaining three also vanish in consequence is quickly seen on closer scrutiny.

A contact net $\mathfrak{C}$ gives rise, in consequence of this theorem, to a unique Möbius net; but the same Möbius net arises from eight different contact nets, corresponding to the eight different ways (*cf.* M.T., p. 349) in which the points of a Möbius configuration can be grouped as a pair of Möbius tetrads.

The derivation of a contact net from a Möbius net.

2. From the equations (1.1) of M.T., and the equations for the three quadrics on p. 341 of that paper, it is seen, on replacing the homogeneous coordinates by their reciprocals, that the quadrics of a contact net are linearly dependent on three quadrics whose equations may be taken to be

$$\left.\begin{aligned} Q_1 &\equiv m(zx+yt)+n(xy-zt) = 0, \\ Q_2 &\equiv n(xy+zt)+l(yz-xt) = 0, \\ Q_3 &\equiv l(yz+xt)+m(zx-yt) = 0. \end{aligned}\right\} \qquad \text{(I)}$$

This is a canonical form for the quadrics of $\mathfrak{C}$. The three quadrics $Q_1 = 0$, $Q_2 = 0$, $Q_3 = 0$ contain each a pair of opposite edges of T. The four quadrics

$$Q_2+Q_3-Q_1 = 0, \quad Q_3+Q_1-Q_2 = 0, \quad Q_1+Q_2-Q_3 = 0, \quad Q_1+Q_2+Q_3 = 0,$$

are cones, their vertices being A, B, C, D respectively; each cone contains the three edges of T which pass through its vertex.

The application of the transformation $\mathfrak{S}$ to the Möbius net affords an instantaneous proof of the theorem. For, the base points of the Möbius net being given by the rows of

$$\begin{pmatrix} . & -n & m & -l \\ n & . & -l & -m \\ -m & l & . & -n \\ l & m & n & . \end{pmatrix},$$

the lines $\alpha, \beta, \gamma, \delta$ which arise on applying the transformation meet the opposite faces of T in the points given by the rows of

$$\begin{pmatrix} . & -n^{-1} & m^{-1} & -l^{-1} \\ n^{-1} & . & -l^{-1} & -m^{-1} \\ -m^{-1} & l^{-1} & . & -n^{-1} \\ l^{-1} & m^{-1} & n^{-1} & . \end{pmatrix},$$

and this matrix is also of the standard form for the Möbius configuration.

A cone of the Möbius net becomes, when subjected to $\mathfrak{S}$, a cone of $\mathfrak{C}$, the vertices of the two cones being corresponding points in the transformation. Thus those cones whose vertices lie on the line e (one of the two transversals, e and f, of AA', BB', CC', DD'; *cf.* M.T., § 6) become cones whose vertices lie on the curve which is the transformation of e, namely a twisted cubic k_1 circumscribing T. And the cones whose vertices lie on f become cones whose vertices lie on k_2, a second twisted cubic circumscribing T. The Jacobian curve of $\mathfrak{C}$ is therefore constituted by k_1 and k_2.

Again: the surface which is the envelope of the cones of the Möbius net whose vertices are on e is transformed into the surface which is the envelope of the cones of $\mathfrak{C}$ whose vertices are on k_1. Now (*cf.* M.T., § 7) the cones of the Möbius net whose vertices lie on e envelop a Plücker surface U having e as a nodal line and nodes at the eight points of the configuration. The transformation by $\mathfrak{S}$ of any quartic surface which has nodes at the vertices of T is another quartic surface likewise having nodes at the vertices of T, so that, in particular, U is transformed into a quartic surface. This quartic surface must, since e is a nodal line on U, have the twisted cubic k_1, which is the transformation of e, for a nodal curve; it is therefore a scroll $R_1{}^4$, and the cones of $\mathfrak{C}$ whose vertices lie on k_1 are the tangent cones to this scroll from the points of its nodal curve. Those cones of $\mathfrak{C}$ whose vertices lie on k_2 likewise envelope a quartic scroll $R_2{}^4$, and are the tangent cones to this scroll from the points

of its nodal curve. The envelope of the cones of © consists of R_1^4 and R_2^4; these two scrolls, being both of class four, make up an envelope of total class eight, which is known to be the correct class for the envelope of the cones which belong to a net of quadrics.

3. Consider now the scroll generated by the trisecants of the Jacobian curve ϑ. Since ϑ here consists of k_1 and k_2, any trisecant must be a chord of one curve and a secant of the other; the question therefore arises, whether the line conjugate to a point of k_1 is a chord of k_1 and a secant of k_2, or whether it is a secant of k_1 and a chord of k_2. This is answered at once by appealing to a known property of ϑ; namely that the trisecants which are conjugate to the vertices of any tetrahedron which form a canonical set on ϑ lie in the opposite faces of this tetrahedron*. For, since (M.T., p. 342) a canonical set on the Jacobian curve of the Möbius net consists of two points on e and two points on f, a canonical set on the Jacobian curve of the contact net must consist of two points on k_1 and two points on k_2. The face of the tetrahedron, whose vertices are such a set of four points, which is opposite to one of the points P on k_1 has two of its three vertices on k_2; thus the remaining three intersections of this face with the Jacobian curve, which lie on the trisecant conjugate to P, can only include one point of k_2. So the trisecants conjugate to the points of k_1 are chords of k_1 and secants of k_2, while those conjugate to the points of k_2 are chords of k_2 and secants of k_1.

The scroll of order eight generated by the trisecants of the Jacobian curve thus consists of the quartic scroll K_1 generated by those chords of k_1 which meet k_2, and of the quartic scroll K_2 generated by those chords of k_2 which meet k_1. Those trisecants which are conjugate to the common points A, B, C, D of k_1 and k_2 are common generators of K_1 and K_2; these are in fact the lines α, β, γ, δ, since the polar plane of A with respect to any quadric of © contains α, and similarly for the other vertices of T. The common curve of K_1 and K_2 consists of the four lines α, β, γ, δ and of the two curves k_1 and k_2, both counted twice, since each curve is nodal on one scroll and simple on the other.

The derivation of a contact net from a quartic scroll.

4. If a surface F has a nodal curve g there are, on this nodal curve, a certain number of *pinch-points*; at any point of g there are two tangent planes to F, both passing through the tangent of g; the pinch-points are

* *Proc. London Math. Soc.* (2), 44 (1938), 467.

those points of g for which these two planes coincide. The tangent plane of F at a pinch-point meets the surface in a curve having a triple point there; of the three branches of the curve, two touch the tangent of g at the pinch-point while the third has another tangent, this latter being called by Cayley* the *cotangent*. The tangent cone to F from any point P (apart from the cone which projects g from P) touches F along the *polar curve*† of P, and this polar curve, whatever the position of P, passes through each pinch-point and touches the cotangent there.

Suppose now that F is a scroll. The two tangent planes of F at a point of its nodal curve are the planes which join the tangent of the nodal curve to the two generators which pass through the point. But, whereas through a general point of the nodal curve there pass two generators, through a pinch-point there passes only one. This is a torsal generator, and is the cotangent at the pinch-point. Hence the polar curve of any point P passes through the pinch-points and has the torsal generators for its tangents there.

When F is a quartic scroll R^4 its nodal curve is a twisted cubic k, and there are four pinch-points A, B, C, D; through these pass the respective torsal generators $\alpha, \beta, \gamma, \delta$. The polar curve of any point P is a sextic, touching $\alpha, \beta, \gamma, \delta$ at A, B, C, D. If, however, P lies on the nodal curve, its polar curve includes the two generators through P; the remaining part is a quartic, having a node at P and touching $\alpha, \beta, \gamma, \delta$ at A, B, C, D. The tangents to R^4 from P are the chords of this curve which pass through P, so that they generate a quadric cone which passes through A, B, C, D and has $\alpha, \beta, \gamma, \delta$ for tangent lines there. That this cone of tangents is of the second order is also readily seen on taking a plane section through P, for there are two tangents to a trinodal plane quartic from one of its nodes.

There has thus arisen a singly-infinite set of quadric cones, namely the tangent cones to R^4 from the points of k, all of which pass through A, B, C, D and have $\alpha, \beta, \gamma, \delta$ as tangent lines there. These cones must belong either to a contact net or to a pencil; in the latter event they would constitute the whole of the pencil. But there does not exist a pencil of quadric cones the locus of whose vertices is a twisted cubic; therefore the cones belong to a contact net $\mathfrak{C}$.

Since a quartic scroll gives rise in this way to a contact net, the theorem may be stated as a property (as yet apparently unnoticed) of a quartic

* *Collected Mathematical Papers*, 6, 335.

† This is the curve called the *pure polar* by Miss Hudson in *Cremona transformations* (Cambridge, 1927), 372, and the *proper curve of contact* by Baker in *Principles of geometry*, 6 (Cambridge, 1933), 150.

scroll:

The four points in which the torsal generators of a quartic scroll meet the faces of the tetrahedron whose vertices are the pinch-points form, when taken with the pinch-points, a pair of Möbius tetrads.

5. It is now evident that *quartic scrolls occur in pairs.* For let a quartic scroll R_1^4 be given. It has a nodal twisted cubic k_1, and the tangent cones to R_1^4 from the points of k_1 belong to a contact net $\mathfrak{C}$. The cones of $\mathfrak{C}$, however, fall into two families, of which this set of tangent cones of R_1^4 is but one; the Jacobian curve of $\mathfrak{C}$ consists of k_1 and of a second twisted cubic k_2 meeting k_1 at the base points of $\mathfrak{C}$ (*i.e.*, the pinch-points of R_1^4). The envelope of this second family of cones is a second quartic scroll R_2^4, which is thus obtainable when R_1^4 is given.

These two scrolls, together constituting the envelope of the cones of $\mathfrak{C}$, are symmetrically related to one another; R_1^4 may be derived from R_2^4 exactly as R_2^4 has been derived from R_1^4. The base points of $\mathfrak{C}$ are pinch-points, and the fixed tangent lines of $\mathfrak{C}$ are torsal generators, *on both scrolls.*

6. Let the nodal curve k of a quartic scroll R^4 be given by

$$x : y : z : t = \theta^3 : \theta^2 : \theta : 1.$$

Then if P is the point on k whose parameter is θ, the equation of any quadric cone whose vertex is P is given by equating to zero some quadratic form in the three variables

$$x-\theta y, \quad y-\theta z, \quad z-\theta t.$$

If this cone is the tangent cone from P to R^4, the coefficients in the quadratic form depend only on R^4 itself, and do not vary with θ; thus the equation of this tangent cone has the form

$$\theta^2 S_0+2\theta S_1+S_2 = 0, \tag{6.1}$$

where S_0, S_1, S_2 are certain quadratic forms, fixed when R^4 is given, in x, y, z, t. These cones clearly belong to a net of quadrics.

So much is clear without using the explicit equation of R^4; but now suppose that this equation is

$$(a, b, c, f, g, h)(zx-y^2, xt-yz, yt-z^2)^2 = 0. \tag{6.2}$$

Then the chord which joins the points θ_1 and θ_2 on k is a generator of R^4 provided that

$$(a, b, c, f, g, h\mathfrak{J}\theta_1\theta_2, \theta_1+\theta_2, 1)^2 = 0. \qquad (6.3)$$

The enveloping cone from P to R^4 is the envelope of those planes which join P to the generators of R^4. If P lies on k and has the parameter θ, the equation of the plane which joins it to the chord joining the points θ_1 and θ_2 of k is

$$x-y(\theta+\theta_1+\theta_2)+z(\theta\theta_1+\theta\theta_2+\theta_1\theta_2)-t\theta\theta_1\theta_2 = 0,$$

which may be written

$$\theta_1\theta_2(z-\theta t)+(\theta_1+\theta_2)(\theta z-y)+x-\theta y = 0.$$

The envelope of this plane, subject to (6.3), is, by clear analogy with the theory of conic sections, the quadric cone

$$(A, B, C, F, G, H\mathfrak{J}z-\theta t, \theta z-y, x-\theta y)^2 = 0,$$

where $A = bc-f^2, \ldots, H = fg-ch$.

This, then, is the equation of the cone of tangent lines from P to R^4; it is in complete agreement with (6.1), as is seen by writing

$$S_0 \equiv (A, B, C, F, G, H\mathfrak{J}-t, z, -y)^2,$$

$$S_1 \equiv (A, B, C, F, G, H\mathfrak{J}-t, z, -y\mathfrak{J}z, -y, x),$$

$$S_2 \equiv (A, B, C, F, G, H\mathfrak{J}z, -y, x)^2.$$

It was remarked by Cayley* that the three quadrics $S_i = 0$ all meet k at the pinch-points; and, since the four points of k whose parameters satisfy

$$(A, B, C, F, G, H\mathfrak{J}1, -\theta, \theta^2)^2 = 0$$

lie on all these quadrics, they must be the four pinch-points.

Incidentally, since R^4 is the envelope of the tangent cones from the points of k, its equation, by (6.1), must be $S_0 S_2 = S_1^2$. With the above forms for S_0, S_1, S_2 this is the same equation as (6.2).

Since every cone (6.1) must touch every generator of R^4, and every cone (6.1) passes through the pinch-points, the torsal generators of R^4 must touch every one of the cones at the pinch-points. This shows that the net of quadrics to which the cones belong is a contact net.

The two points θ and ϕ of k are found to be conjugate with respect to every cone (6.1), and therefore with respect to every quadric of the

* *Collected Mathematical Papers*, 6, 124.

contact net to which these cones belong, provided only that

$$(A, B, C, F, G, H\overline{\times}1, -\theta, \theta^2\overline{\times}1, -\phi, \phi^2) = 0. \qquad (6.4)$$

Hence the polar planes, of any point ϕ of k, with respect to the tangent cones to R^4 from all the different points of k, have in common a line which is a chord of k. This is in accordance with the result previously found for $\mathcal{C}$, that the trisecant conjugate to a point of either twisted cubic which is part of the Jacobian curve is a chord of this cubic (and a secant of the other).

If θ_1 and θ_2 are the two roots of (6.4) for any given value of ϕ, then

$$\theta_1\theta_2 : \theta_1+\theta_2 : 1 = A-H\phi+G\phi^2 : H-B\phi+F\phi^2 : G-F\phi+C\phi^2,$$

so that, whatever the value of ϕ,

$$\{a\theta_1\theta_2+h(\theta_1+\theta_2)+g\}\{g\theta_1\theta_2+f(\theta_1+\theta_2)+c\} = \{h\theta_1\theta_2+b(\theta_1+\theta_2)+f\}^2.$$

This is of the same form as (6.3), with the constants

$$a, \quad b, \quad c, \quad f, \quad g, \quad h$$

of (6.3) replaced respectively by

$$ag-h^2, \ hf-b^2, \ cg-f^2, \ \tfrac{1}{2}(ch+fg)-bf, \ \tfrac{1}{2}(ac+g^2)-hf, \ \tfrac{1}{2}(af+gh)-bh.$$

If this replacement is carried out in (6.2), there results the equation of the quartic scroll K generated by those chords of k which are conjugate to the points of k. The chord which is conjugate to a pinch-point of R^4 is the corresponding torsal generator, so that the four torsal generators of R^4 are also (non-torsal) generators of K.

Mathematical Institute,
16 Chambers Street,
Edinburgh, 1.

NOTE ON COMBINATORIAL ANALYSIS

By R. RADO.

[Received 16 September, 1939.—Read 15 December, 1939.]

Consider a system of equations

$$a_{\mu 1}x_1+a_{\mu 2}x_2+\ldots+a_{\mu n}x_n=b_\mu \quad (1\leqslant\mu\leqslant m), \tag{1}$$

where the $a_{\mu\nu}$, b_μ are complex numbers. Let A be a set of numbers, for instance all numbers of a number field, all complex numbers different from zero, etc. We call (1) *regular in* A if the following condition holds: however we split A into a finite number of subsets $A_1, A_2, \ldots, A_k$, always at least one of these subsets A_κ contains a solution of (1). Roughly speaking, regularity of (1) in A means that, in a certain sense, A contains very many solutions of (1), and these solutions interlock very intimately.

In the special case where A is the set of all positive integers, I. Schur† proved the regularity of

$$x_1+x_2-x_3=0,$$

and van der Waerden‡ proved the regularity, for every $l>0$, of

$$x_0-x_1=x_1-x_2=\ldots=x_{l-1}-x_l\neq 0\text{§}.$$

In a previous note‖ I determined all systems (1) which are regular in the set of all positive integers. In this case necessary and sufficient conditions for regularity turned out to be certain linear relations between the $a_{\mu\nu}$, b_μ. In the present note, which is an elaboration of a lecture delivered at the International Congress of Mathematicians at Oslo, 1936¶,

† *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 25 (1916), 114

‡ *Nieuw Archief voor Wiskunde*, 15 (1927), 212–216.

§ Regularity of a system of equations and inequalities, in fact, of any set of conditions imposed upon certain variables, is defined in exactly the same way as above,

‖ *Math. Zeitschrift*, 36 (1933), 424–480, quoted as S,

¶ *Comptes Rendus, Oslo*, 2 (1936), 20–21,

I propose to consider the same problem in the case of more general sets A. I establish necessary and sufficient conditions for regularity expressed in terms of linear relations between the $a_{\mu\nu}$, b_μ in the following cases:

(i) The $a_{\mu\nu}$ are arbitrary, $b_\mu = 0$, A is the set of all numbers different from zero contained in a given ring of complex numbers.

(ii) The $a_{\mu\nu}$ and b_μ are arbitrary numbers, A is the field of all algebraic numbers.

(iii) The $a_{\mu\nu}$ are algebraic numbers, the b_μ are arbitrary, A is the field of all complex numbers.

Other cases can be dealt with which are not included in this note.

The criteria are analogous to those obtained in the special case of S. In the cases (ii), (iii) the condition for regularity postulates that, for some number ξ of A,

$$\sum_{\nu=1}^{n} a_{\mu\nu}\xi = b_\mu \quad (1 \leqslant \mu \leqslant m).$$

This result may be expressed as follows. *If, in every distribution of the numbers of A over a finite number of classes, at least one class contains a solution of* (1), *then, in cases* (ii) *and* (iii), *the same is true for the extreme case of a distribution in which every number of A forms a class by itself.*

Some of the proofs are extensions of proofs in S, others require the use of different methods. In proving the result concerning (i), we employ an extension (Theorem II) of van der Waerden's theorem quoted above. This extension was first proved by Dr. G. Grünwald, who kindly communicated it to me. It may be stated as follows. *Given any "configuration" S consisting of a finite number of lattice points† of a Euclidean space, and given a distribution of all lattice points of this space into a finite number of classes, there is at least one class which contains a configuration S' of lattice points which is similar and parallel (homothetic) to S.* Dr. Grünwald's proof runs parallel to van der Waerden's proof of his theorem. The proof given in this note is a simplification analogous to the simplification of van der Waerden's proof given in S (p. 432, Satz I). In an earlier note‡ I proved a weaker form of Grünwald's Theorem in which similarity of S and S' was replaced by affinity.

The last paragraph deals with regularity of systems (1) with respect to distributions which have denumerably many classes.

---

† *I.e.* points with rational integral coordinates.

‡ *Berliner Sitzungsberichte* (1933), 589–596, Satz I.

1. *Preliminaries. Generalisation of van der Waerden's Theorem.*

1. Let A be a finite or infinite aggregate. In this section letters a, b, c, x denote general elements of A. Throughout this paper the letter Δ (and Δ', Δ'', ..., Δ_1, Δ_2, etc.) denotes distributions of all elements of A into a finite number of classes. Occasionally we consider distributions into an infinite number of classes, in which case this is mentioned explicitly. Δ is defined by means of a relation "$\sim$" which is defined for some pairs of elements of A and which has the properties:

(i) $a \sim a$ for every a;

(ii) $a \sim b$ implies $b \sim a$;

(iii) $a \sim b$ and $b \sim c$ imply $a \sim c$;

(iv) every infinite subset of A contains two distinct elements a, b such that $a \sim b$.

By $|\Delta|$ we denote the number of non-empty classes of Δ. A congruence

$$a \equiv b \pmod{\Delta}$$

expresses, by definition, the fact that a and b belong to the objects distributed by means of Δ, and, moreover, belong to the same class. In analogy with the notation for functions, we speak of a distribution $\Delta(x)$ defined for every x of A, or, briefly, defined in A.

Throughout this paper $\Delta^{(k)}(x)$, for every positive integer k, denotes the distribution of all rational integers into classes of equal residues mod k. Thus

$$x \equiv y \pmod{\Delta^{(k)}}$$

is equivalent to saying that x and y are rational integers and

$$x \equiv y \pmod{k}.$$

$\Delta^{(0)}$ denotes that distribution of A in which every element of A forms a class for itself. $\Delta^{(0)}$ may have infinitely many classes. We use the same symbol $\Delta^{(0)}$ for different sets A.

Two methods are employed for generating new distributions from given ones†. The first is a process of multiplication. Given a finite

† Both were used in S. The notation adopted in this paper seems to be more convenient than the one used in S.

number of distributions $\Delta_1, \Delta_2, \ldots, \Delta_n$, each defined in A, we understand by their *product*

$$\Delta = \Delta_1 \Delta_2 \ldots \Delta_n = \prod_{\nu=1}^{n} \Delta_\nu$$

that distribution Δ of A which is defined by the rule:

$$x \equiv y \pmod{\Delta}$$

if, and only if,

$$x \equiv y \pmod{\Delta_\nu} \quad (1 \leqslant \nu \leqslant n).$$

For instance, if k and l are natural numbers, then

$$\Delta^{(k)} \Delta^{(l)} = \Delta^{(m)},$$

where m is the least common multiple of k and l.

We have

$$|\Delta_1 \Delta_2 \ldots \Delta_n| \leqslant |\Delta_1||\Delta_2| \ldots |\Delta_n|.$$

The second process is one of *inducing a distribution* in a set B by means of a distribution in A and a correspondence between every element of B and some elements of A. Suppose that $\Delta(x)$ is defined in A, and that $f(y)$ is a function defined for every element y of a set B. The functional values of f are elements of A. Then we define a distribution $\Delta_1(y)$ in B by postulating that

$$y_1 \equiv y_2 \pmod{\Delta_1}$$

is to be equivalent to

$$f(y_1) \equiv f(y_2) \pmod{\Delta}.$$

We use the notation

$$\Delta_1(y) = \Delta\big(f(y)\big).$$

Clearly

$$|\Delta_1| \leqslant |\Delta|.$$

We have, for instance,

$$\Delta^{(k)}(x+1) = \Delta^{(2k)}(2x) = \Delta^{(k)}(x).$$

For

$$x_1 \equiv x_2 \pmod{k}$$

is equivalent to

$$x_1 + 1 \equiv x_2 + 1 \pmod{k}$$

and also to

$$2x_1 \equiv 2x_2 \pmod{2k}.$$

For real numbers $x \neq 0$,

$$\Delta^{(0)}\left(\frac{x}{|x|}\right) = \Delta^{(3)}\left(\frac{x}{|x|}\right)$$

is a distribution in which all positive numbers are in one class and all negative numbers are in a second class.

Suppose that $\Delta(x)$ is defined in A and that the classes of Δ are the sets

$$A_0, A_1, \ldots, A_{k-1}.$$

Define, for every x of A, a function $f(x)$ by means of†

$$x \prec A_{f(x)}.$$

Then
$$\Delta(x) = \Delta^{(k)}\big(f(x)\big).$$

Now choose a natural number l. Then corresponding to every x of A there are integers $g(x)$, $h(x)$ such that

$$f(x) = g(x) + lh(x),$$

$$0 \leqslant g(x) < l; \quad 0 \leqslant h(x) \leqslant (k-1)/l.$$

Therefore

$$\Delta(x) = \Delta^{(k)}\big(f(x)\big) = \Delta^{(0)}\big(g(x)\big)\,\Delta^{(k)}\big(h(x)\big) = \Delta'(x)\,\Delta''(x),$$

say. We have

$$|\Delta'| \leqslant l, \quad |\Delta''| \leqslant \left[\frac{k-1}{l}\right] + 1‡.$$

Hence every Δ can be represented in the form

$$\Delta = \Delta'\Delta'',$$

where
$$|\Delta'| \leqslant l, \quad |\Delta''| \leqslant \left[\frac{|\Delta|-1}{l}\right] + 1.$$

In particular, putting $l = 2$, we see that every Δ is a product of a finite number of distributions with not more than two classes.

If x, y are general elements of two sets A, B respectively, then $\Delta(x, y)$ denotes a distribution of all pairs (x, y). Thus, for rational integers x, y,

$$\Delta(x, y) = \Delta^{(k)}(2x + 3y)$$

---

† We use the symbol "$\prec$" to denote the relation of an element to the class to which it belongs.

‡ $[t]$ denotes the largest integer not exceeding t.

denotes that distribution of all pairs (x, y) for which (x_1, y_1) and (x_2, y_2) are in the same class if, and only if,

$$2x_1+3y_1 \equiv 2x_2+3y_2 \pmod{k}.$$

2. Let $S(x_1, x_2, \ldots, x_n)$ be a system of conditions imposed upon the values of variables $x_1, \ldots, x_n$. These variables are allowed to vary throughout a set A. A relation

$$S(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) = 0 \tag{2}$$

expresses the fact that the elements $x_1^{(0)}, \ldots, x_n^{(0)}$ of A satisfy the conditions (2); and

$$S(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) \neq 0$$

denotes the logical opposite to (2). S is called *k-regular in A* if, however A is split into k subsets

$$A = A_1+A_2+\ldots+A_k,$$

there is always at least one subset A_κ which contains a solution of (2). *Regularity* of S, as mentioned in the introduction, means that S is k-regular for every $k = 1, 2, \ldots$. *ω-regularity* of S means solubility in at least one class whenever A is split into denumerably many classes. Finally, we call S *absolutely regular in A* if, for some $x^{(0)}$ of A,

$$S(x^{(0)}, x^{(0)}, \ldots, x^{(0)}) = 0.$$

If A is the set of all real numbers except zero the condition

$$(x_1-x_2-1)(x_1-x_2-2)\ldots(x_1-x_2-k) = 0$$

is k-regular, but not $(k+1)$-regular, the condition

$$x_1+x_2-x_3 = 0$$

is regular but not ω-regular†, and the condition

$$x_1 \neq x_2$$

is ω-regular but not absolutely regular in A. The *degree of regularity* of S in A is the largest natural number k (if there is one) such that S is k-regular in A.

Let $S'(x_1, x_2, \ldots, x_{n'})$ be a system of conditions with $n' \leqslant n$ which has the property that (2) implies

$$S'(x_1^{(0)}, x_2^{(0)}, \ldots, x_{n'}^{(0)}) = 0.$$

† See below, Theorems VII and XI.

Let k, k' be natural numbers, $k' \leqslant k$, and let A be a subset of A'. Then k-regularity of S in A implies k'-regularity of S' in A'. Theorem I states that in certain circumstances a kind of inverse of this implication holds. In fact, many of our results are of this type.

THEOREM I. *If $S(x_1, x_2, \ldots, x_n)$ is k-regular in a denumerable set A then S is also k-regular in a suitable finite subset A' of A.*

Theorem I has been proved elsewhere†, but, for convenience, I reproduce the proof.

Proof. We may suppose that

$$A = \{1, 2, 3, \ldots\}.$$

Put $$A_N = \{1, 2, \ldots, N\} \qquad (N = 1, 2, 3, \ldots).$$

Let us assume that, for no value of N, $S(x_1, \ldots, x_n)$ is k-regular in A_N. We have to show that S is not k-regular in A. There is a distribution Δ_N of A_N,

$$A_N = A_{N1} + A_{N2} + \ldots + A_{Nk},$$

such that no set $A_{N\kappa}$ contains a solution of $S = 0$. Here $N = 1, 2, \ldots$; $1 \leqslant \kappa \leqslant k$. For any positive integer x we denote by $f_N(x)$ that number κ $(1 \leqslant \kappa \leqslant k)$ for which

$$x \prec A_{N\kappa}.$$

Then by a well-known argument (Cantor's "Diagonalverfahren") we can define a function $f^*(x)$ such that, given any $x_0 > 0$, there are infinitely many N's for which

$$f_N(x) = f^*(x) \quad (1 \leqslant x \leqslant x_0). \tag{3}$$

Let, for $1 \leqslant \kappa \leqslant k$, A_κ^* be the set of all x for which

$$f^*(x) = \kappa,$$

and let Δ^* be the distribution

$$A = A_1^* + A_2^* + \ldots + A_k^*.$$

Now consider n numbers x_ν' satisfying

$$x_1' \equiv x_2' \equiv \ldots \equiv x_n' \pmod{\Delta^*} \tag{4}$$

† S., Satz III. See also Dénes König, *Theorie der endlichen und unendlichen Graphen* (Leipzig, 1936), 84 (α) and (β).

Put $$x_0 = \max(x_1', x_2', \ldots, x_n').$$

Then, by the definition of $f^*(x)$, there is a number $N > x_0$ for which (3) holds. In particular, using (4), we have

$$f_N(x_\nu') = f^*(x_\nu') = \kappa_0 \quad (1 \leqslant \nu \leqslant n),$$

$$x_\nu' \prec A_{N\kappa_0} \qquad (1 \leqslant \nu \leqslant n).$$

Hence, by the definition of Δ_N,

$$S(x_1', x_2', \ldots, x_n') \neq 0.$$

This completes the proof of Theorem I.

As a corollary of Theorem I we prove†

LEMMA 1. *Let B be a denumerable set of numbers. Let R be the ring and K be the field generated by B. Let $R-\{0\}$ and $K-\{0\}$ denote the sets of all non-zero numbers of R and K respectively. Then k-regularity of the system*

$$\sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m) \tag{5}$$

in $K-\{0\}$ implies k-regularity of the same system in $R-\{0\}$.

This lemma is another kind of inverse of the simple proposition on p. 128, immediately preceding Theorem I.

Proof of Lemma 1. Suppose that (5) is k-regular in $K-\{0\}$. Then, by Theorem I, since K is denumerable, (5) is k-regular in a finite subset T of $K-\{0\}$. Let N be the product of the "denominators" of all numbers of T, or, more accurately, let N be a number of $R-\{0\}$ for which

$$Nt \prec R-\{0\} \tag{6}$$

whenever $t \prec T$. Take any $\Delta(x)$ defined in $R-\{0\}$ and satisfying $|\Delta| \leqslant k$. Put

$$\Delta'(x) = \Delta(Nx) \quad (x \prec T).$$

In view of (6) this definition is significant. From $|\Delta'| \leqslant |\Delta| \leqslant k$, and from the definition of T, we deduce the existence of numbers t_ν of T for which

$$\sum_{\nu=1}^{n} a_{\mu\nu} t_\nu = 0 \qquad (1 \leqslant \mu \leqslant m),$$

$$t_1 \equiv t_2 \equiv \ldots \equiv t_n \qquad (\text{mod } \Delta').$$

† S, 441.

Then $$\sum_{\nu} a_{\mu\nu}(Nt_\nu) = 0 \qquad (1 \leqslant \mu \leqslant m),$$

$$Nt_1 \equiv Nt_2 \equiv \ldots \equiv Nt_n \pmod{\Delta}.$$

Hence the result. Instead of (5) we might have taken any system of homogeneous conditions.

3. In this section small Latin letters denote non-negative integers. Capitals $A, B, \ldots, X$ denote "vectors" $(x_1, x_2, \ldots, x_m)$ of a fixed dimension m. If

$$A = (x_1, x_2, \ldots, x_m), \quad A' = (x_1', x_2', \ldots, x_m'),$$

we put

$$|A| = \max(x_1, x_2, \ldots, x_m), \quad aA = (ax_1, \ldots, ax_m),$$

$$A + A' = (x_1 + x_1', x_2 + x_2', \ldots, x_m + x_m').$$

0 denotes the number zero as well as the vector $(0, \ldots, 0)$.

THEOREM II. *There is a function $f(k, R_1, R_2, \ldots, R_l)$ which has the following property. Suppose that $R_1, R_2, \ldots, R_l$ are vectors and that Δ is a distribution of all vectors into k classes. Then we can find A^*, d^*, $(d^* > 0)$ such that*

(7) $$A^* + d^* R_\lambda \equiv A^* \pmod{\Delta} \quad (1 \leqslant \lambda \leqslant l),$$

8) $$|A^* + d^* R_\lambda| \leqslant f \quad (1 \leqslant \lambda \leqslant l).$$

Proof. The theorem is true for $l = 1$. For there are two of the $k+1$ vectors

$$0, R_1, 2R_1, 3R_1, \ldots, kR_1$$

which belong to the same class of Δ, say αR_1 and βR_1, where

$$0 \leqslant \alpha < \beta \leqslant k.$$

Put $$A^* = \alpha R_1, \quad d^* = \beta - \alpha.$$

Then $$A^* + d^* R_1 = \beta R_1 \equiv \alpha R_1 = A^* \pmod{\Delta}.$$

Since $$|A^* + d^* R_1| = \beta |R_1| \leqslant k |R_1|,$$

we may put $$f(k, R_1) = k |R_1|.$$

Therefore we may suppose that, for some given vectors

$$R_1, R_2, \ldots, R_l$$

$(l > 1)$, the existence of $f(k', R_1, R_2, \ldots, R_{l-1})$ has been established for every $k' > 0$, and we deduce the existence of

$$f(k, R_1, \ldots, R_{l-1}, R_l).$$

In the proof which follows, $k, R_1, \ldots, R_l$ are constant, and these numbers are not shown in the arguments of functions. Δ is a distribution of all vectors, $|\Delta| \leqslant k$. We may assume, without loss of generality, that $R_1 \neq 0$.

LEMMA 2. *Under the assumptions stated there is a function $g(n)$ such that, given any A, n, we can find A', d' $(d' > 0)$ such that*

$$A+(A'+d' R_\lambda)+B \equiv A+A'+B \pmod{\Delta} \quad (1 \leqslant \lambda < l,\ |B| \leqslant n), \tag{9}$$

$$|A'+d' R_\lambda| \leqslant g(n) \quad (1 \leqslant \lambda < l). \tag{10}$$

Proof. Put

$$\Delta'(X) = \prod_{|B| \leqslant n} \Delta(A+X+B).$$

Then

$$|\Delta'| \leqslant g_1(n).$$

Using the definition of

$$f\big(g_1(n), R_1, R_2, \ldots, R_{l-1}\big) = g_2(n),$$

we find that there are an A' and d' $(d' > 0)$ such that

$$A'+d' R_\lambda \equiv A' \pmod{\Delta'} \quad (1 \leqslant \lambda < l), \tag{11}$$

$$|A'+d' R_\lambda| \leqslant g_2(n) \quad (1 \leqslant \lambda < l). \tag{12}$$

(11) and (12) are equivalent to (9) and (10) respectively if we put

$$g(n) = g_2(n).$$

Thus the lemma is proved. We note that

$$|A'+d' R_l| \leqslant |A'+d' R_1| + |A'+d' R_1||R_l| \leqslant g(n)+g(n)|R_l| = h(n).$$

Now, to prove Theorem II, put

$$n_k = 1, \quad n_\kappa = \sum_{\nu=\kappa+1}^{k} h(n_\nu) \quad (0 \leqslant \kappa < k).$$

Apply Lemma 2, with $A = 0$, $n = n_0$. We find

$$A' = A_0, \quad d' = d_0 > 0$$

such that

$$0+(A_0+d_0 R_\lambda)+B \equiv 0+A_0+B \pmod{\Delta} \quad (1 \leqslant \lambda < l;\ |B| \leqslant n_0),$$
$$|A_0+d_0 R_\lambda| \leqslant h(n_0) \quad (1 \leqslant \lambda \leqslant l).$$

A second application, with $A = A_0$, $n = n_1$, yields $A' = A_1$, $d' = d_1 > 0$, such that

$$A_0+(A_1+d_1 R_\lambda)+B \equiv A_0+A_1+B \pmod{\Delta} \quad (1 \leqslant \lambda < l;\ |B| \leqslant n_1),$$
$$|A_1+d_1 R_\lambda| \leqslant h(n_1) \qquad (1 \leqslant \lambda \leqslant l).$$

Proceeding in this way (the next case is $A = A_0+A_1$, $n = n_2$) we find $A' = A_\kappa$, $d' = d_\kappa > 0$ such that

$$\sum_{\nu=0}^{\kappa-1} A_\nu+(A_\kappa+d_\kappa R_\lambda)+B \equiv \sum_{\nu=0}^{\kappa-1} A_\nu+A_\kappa+B \pmod{\Delta}\dagger \tag{13}$$
$$(0 \leqslant \kappa \leqslant k;\ 1 \leqslant \lambda < l;\ |B| \leqslant n_\kappa),$$
$$|A_\kappa+d_\kappa R_\lambda| \leqslant h(n_\kappa) \quad (0 \leqslant \kappa \leqslant k;\ 1 \leqslant \lambda \leqslant l). \tag{14}$$

There are two among the $k+1$ vectors

$$V_\kappa = \sum_{\nu=0}^{\kappa} A_\nu + \sum_{\nu=\kappa+1}^{k} (A_\nu+d_\nu R_l) \quad (0 \leqslant \kappa \leqslant k)$$

which belong to the same class of Δ. Therefore, say,

$$V_\alpha \equiv V_\beta \pmod{\Delta},$$

where $0 \leqslant \alpha < \beta \leqslant k$. Put

$$A^* = V_\beta, \quad d^* = \sum_{\nu=\alpha+1}^{\beta} d_\nu.$$

Then

$$A^*+d^* R_l = \sum_{0}^{\beta} A_\nu + \sum_{\beta+1}^{k} (A_\nu+d_\nu R_l) + \sum_{\alpha+1}^{\beta} d_\nu R_l \tag{15}$$
$$= \sum_{0}^{\alpha} A_\nu + \sum_{\alpha+1}^{k} (A_\nu+d_\nu R_l) = V_\alpha \equiv V_\beta = A^* \pmod{\Delta}.$$

On the other hand, if $1 \leqslant \lambda < l$, then

$$A^*+d^* R_\lambda = \sum_{0}^{\alpha} A_\nu + (A_{\alpha+1}+d_{\alpha+1} R_\lambda) \tag{16}$$
$$+\left(\sum_{\alpha+2}^{\beta} (A_\nu+d_\nu R_\lambda) + \sum_{\beta+1}^{k} (A_\nu+d_\nu R_l)\right).$$

† Empty sums have the value zero.

Now we have, by (14),

$$\left|\sum_{\alpha+2}^{\beta}(A_\nu+d_\nu R_\lambda)+\sum_{\beta+1}^{k}(A_\nu+d_\nu R_l)\right|\leqslant\sum_{\alpha+2}^{k}h(n_\nu)=n_{\alpha+1}.$$

Therefore, from (13),

$$A^*+d^*R_\lambda\equiv\sum_{0}^{\alpha}A_\nu+A_{\alpha+1}+\left(\sum_{\alpha+2}^{\beta}(A_\nu+d_\nu R_\lambda)+\sum_{\beta+1}^{k}(A_\nu+d_\nu R_l)\right)\pmod{\Delta}. \tag{17}$$

The vector on the right-hand side of (17) is the same as the vector on the right-hand side of (16), except that α is to be replaced by $\alpha+1$. When $\alpha+1<\beta$ a second application of (13) leads to

$$A^*+d^*R_\lambda\equiv\sum_{0}^{\alpha+2}A_\nu+\sum_{\alpha+3}^{\beta}(A_\nu+d_\nu R_\lambda)+\sum_{\beta+1}^{k}(A_\nu+d_\nu R_l)\pmod{\Delta},$$

and so on. After $\beta-\alpha$ steps we find that

$$A^*+d^*R_\lambda\equiv\sum_{0}^{\beta}A_\nu+\sum_{\beta+1}^{h}(A+d_\nu R_l)=V_\beta=A^*\pmod{\Delta}\quad(1\leqslant\lambda<l).$$

This result, together with (15), shows that (7) is true.

Furthermore, by (14), we have, for $1\leqslant\lambda\leqslant l$,

$$|A^*+d^*R_\lambda|=\left|\sum_{0}^{\alpha}A_\nu+\sum_{\alpha+1}^{\beta}(A_\nu+d_\nu R_\lambda)+\sum_{\beta+1}^{k}(A_\nu+d_\nu R_l)\right|\leqslant\sum_{0}^{k}h(n_\nu).$$

Therefore (8) holds, with

$$f(k,\ R_1,\ R_2,\ \ldots,\ R_l)=\sum_{0}^{k}h(n_\nu),$$

and the theorem is proved.

4. THEOREM III. *There is a function $\bar{f}(k, l)$, defined for all pairs (k, l) of positive integers, which has the following property. Suppose that M is a set of objects among which a commutative and associative addition is defined. Let $R_1, R_2, \ldots, R_l$ be elements of M, and let Δ be a distribution of M into k classes. Then there is a positive integer d^* and an element A^* of M such that*

$$A^*+d^*R_\lambda\equiv A^*\pmod{\Delta}\quad(1\leqslant\lambda\leqslant l)\dagger, \tag{18}$$

$$A^*=\sum_{\lambda=1}^{l}a_\lambda^*R_\lambda, \tag{19}$$

$$a_\lambda^*+d^*\leqslant\bar{f}(k, l)\quad(1\leqslant\lambda\leqslant l). \tag{20}$$

† $d^*R_\lambda=R_\lambda+R_\lambda+\ldots+R_\lambda$ (d^* terms).

Proof. Put, for all integers $x_\lambda \geqslant 0$,

$$\Delta'(x_1, x_2, \ldots, x_l) = \Delta\big((x_1+1)R_1+(x_2+1)R_2+\ldots+(x_l+1)R_l\big). \tag{21}$$

Apply Theorem II to Δ' and the "unit vectors"

$$\bar{R}_1 = (1, 0, \ldots, 0),$$
$$\bar{R}_2 = (0, 1, \ldots, 0),$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots$$
$$\bar{R}_l = (0, 0, \ldots, 1).$$

We obtain a vector $(a_1, a_2, \ldots, a_l)$, with integral components $a_\nu \geqslant 0$, and a positive integer d such that

$$(a_1, \ldots, a_l)+d\bar{R}_\lambda \equiv (a_1, \ldots, a_l) \pmod{\Delta'} \quad (1 \leqslant \lambda \leqslant l), \tag{22}$$

$$|(a_1, \ldots, a_l)+d\bar{R}_\lambda| \leqslant f(k, \bar{R}_1, \ldots, \bar{R}_l) = \bar{g}(k, l) \quad (1 \leqslant \lambda \leqslant l), \tag{23}$$

say. (22), in view of (21), is the same as

$$\sum_{\nu=1}^{l}(a_\nu+1)R_\nu + dR_\lambda \equiv \sum_{\nu=1}^{l}(a_\nu+1)R_\nu \pmod{\Delta} \quad (1 \leqslant \lambda \leqslant l).$$

(23) is the same as

$$a_\lambda + d \leqslant \bar{g}(k, l) \quad (1 \leqslant \lambda \leqslant l).$$

Hence we may put

$$A^* = \sum_{\nu=1}^{l}(a_\nu+1)R_\nu, \quad d^* = d; \quad a_\nu^* = a_\nu+1 \ (1 \leqslant \nu \leqslant l),$$
$$\bar{f}(k, l) = \bar{g}(k, l)+1,$$

and the theorem is proved. In the special case where M is the system of vectors

$$A = (x_1, x_2, \ldots, x_m)$$

whose components x_n are non-negative integers, we deduce from (19) and (20) that, in the notation of Theorem II,

$$|A^*+d^*R_\lambda| = \Big|\sum_{\nu=1}^{l} a_\nu^* R_\nu + d^* R_\lambda\Big| \leqslant \sum_\nu (a_\nu^*+d^*)|R_\nu|$$
$$\leqslant \bar{f}(k, l)\sum_\nu |R_\nu| \quad (1 \leqslant \lambda \leqslant l).$$

Therefore the function $f(k, R_1, \ldots, R_l)$ of Theorem II may be assumed to be of the form

$$f(k, R_1, \ldots, R_l) = \bar{f}(k, l)\sum_{\nu=1}^{l}|R_\nu|.$$

For a later application we note that in Theorem III we may stipulate that the elements

$$A^*,\ A^*+d^*R_\lambda \quad (1\leqslant\lambda\leqslant l)$$

are either different from any one out of a finite number of given elements $C_1, C_2, \ldots, C_s$, or are all equal to one of the elements C_σ. In this case (20) must be replaced by

$$a_\lambda^*+d^*\leqslant\bar{f}(k+s,\ l). \tag{24}$$

For suppose that Δ is a distribution of M, $|\Delta|\leqslant k$. Define Δ' by means of

$$A\equiv B \pmod{\Delta'}$$

if, and only if, either

$$A\neq C_\sigma;\quad B\neq C_\sigma\ (1\leqslant\sigma\leqslant s);\quad A\equiv B \pmod{\Delta},$$

or $A=B=C_{\sigma_0}$ for some σ_0, $1\leqslant\sigma_0\leqslant s$. In other words, we remove the C_σ's from their classes in Δ and put them in separate classes. Now apply Theorem III to Δ'. We find A^*, d^*, with $d^*>0$, such that

$$A^*+d^*R_\lambda\equiv A^* \pmod{\Delta'} \quad (1\leqslant\lambda\leqslant l). \tag{25}$$

Furthermore, (19) and (24) hold; (25) implies (18), and, in view of the definition of Δ', the additional condition is satisfied.

2. *Regularity of homogeneous linear equations.*

1. THEOREM IV. *Let $a_1, a_2, \ldots, a_n$ be real numbers. Suppose that no non-empty subset of these numbers has the sum zero. Then the equation*

$$a_1x_1+a_2x_2+\ldots+a_nx_n=0 \tag{26}$$

is not regular in the set of all positive numbers. In particular, if

$$|a_{\nu_1}+a_{\nu_2}+\ldots+a_{\nu_k}|\geqslant(|a_1|+|a_2|+\ldots+|a_n|)/a>0 \tag{27}$$

whenever $\quad 1\leqslant k\leqslant n,\quad 1\leqslant\nu_1<\nu_2<\ldots<\nu_k\leqslant n,$

then the degree of regularity D of (26) in the set of all positive numbers satisfies

$$D<1+\frac{\log b}{\log(1+a^{-1}-b^{-1})}, \tag{28}$$

where b is any arbitrary number exceeding a. In particular,

$$D<1+\tfrac{8}{3}a\log(2a) \quad (\textit{for every } a), \tag{29}$$

$$D<a\log a+o(a\log a) \quad (\textit{as } a\to\infty). \tag{30}$$

If D' is the degree of regularity of (26) *in the set of all real numbers* $x \neq 0$, *then*

$$D' \leqslant 2D+1. \tag{31}$$

Proof. (27) implies $a \geqslant 1$. If no non-empty subset of $a_1, a_2, \ldots, a_n$ has the sum zero, then (27) holds for some $a \geqslant 1$. Therefore the first assertion of the theorem follows from (28). (29) follows easily from (28) by putting $b = 2a$ and making use of the inequality

$$\log(1+t) = t(1-\tfrac{1}{2}t)+\tfrac{1}{3}t^3(1-\tfrac{3}{4}t)+\ldots > t(1-\tfrac{1}{4}), \tag{32}$$

valid for $0 < t \leqslant \frac{1}{2}$.

In order to deduce (30) from (28) we put

$$b = a \log a,$$

for $a > e$. Then (28) becomes

$$D < 1+\frac{\log(a \log a)}{\log\{1+a^{-1}-(a \log a)^{-1}\}} = a \log a+o(a \log a)$$

as $a \to \infty$. If $D < \infty$, then there is a distribution Δ of all positive numbers such that $|\Delta| = D+1$ and

$$a_1 x_1+\ldots+a_n x_n \neq 0$$

whenever $$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta}.$$

Put $$\Delta'(x) = \Delta(|x|)\,\Delta^{(3)}\left(\frac{x}{|x|}\right) \quad (x \text{ real}, \neq 0).$$

In other words, corresponding to every class of Δ we form a new class containing the same numbers but multiplied by -1. Then no class of Δ' contains a solution of (26), and (31) follows.

All that remains to be proved is (28), under the assumption (27). Suppose that (26) is m-regular in the set of all positive numbers, for some positive integer m. Choose a number $q > 1$ and put

$$\Delta(x) = \Delta^{(m)}\left(\left[\frac{\log x}{\log q}\right]\right) \quad (x > 0).$$

Then $|\Delta| = m$. Hence, in consequence of our assumption about m, there are numbers x_ν satisfying

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta},$$

$$a_1 x_1+\ldots+a_n x_n = 0,$$

Put $$\left[\frac{\log x_\nu}{\log q}\right] = m_\nu \quad (1 \leqslant \nu \leqslant n).$$

Then

$$m_1 \equiv m_2 \equiv \ldots \equiv m_n \pmod{m}, \tag{33}$$

$$q^{m_\nu} \leqslant x_\nu < q^{m_\nu+1} \quad (1 \leqslant \nu \leqslant n). \tag{34}$$

Arrange the m_ν in non-increasing order:

$$m_{\nu_1} \geqslant m_{\nu_2} \geqslant \ldots \geqslant m_{\nu_n}.$$

We have, for some suitable k $(1 \leqslant k \leqslant n)$†,

$$m_{\nu_1} = m_{\nu_2} = \ldots = m_{\nu_k} > m_{\nu_{k+1}}.$$

Then, by (33), $$m_{\nu_k} \geqslant m_{\nu_{k+1}} + m.$$

Therefore, from (34),

$$q^{m_{\nu_1}} \leqslant x_{\nu_\kappa} < q^{m_{\nu_1}+1} \qquad (1 \leqslant \kappa \leqslant k),$$

$$x_{\nu_\lambda} < q^{m_{\nu_\lambda}+1} \leqslant q^{m_{\nu_1}-m+1} \quad (k < \lambda \leqslant n).$$

Now, making use of (27), we deduce that

$$0 = \left| \sum_{\mu=1}^{n} a_\mu x_\mu \right| = \left| \sum_{\kappa=1}^{k} a_{\nu_\kappa} x_{\nu_1} + \sum_{\kappa=1}^{k} a_{\nu_\kappa}(x_{\nu_\kappa} - x_{\nu_1}) + \sum_{\lambda=k+1}^{n} a_{\nu_\lambda} x_{\nu_\lambda} \right|$$

$$\geqslant x_{\nu_1} \left| \sum_{\kappa=1}^{k} a_{\nu_\kappa} \right| - \sum_{\kappa=1}^{k} |a_{\nu_\kappa}||x_{\nu_\kappa} - x_{\nu_1}| - \sum_{\lambda=k+1}^{n} |a_{\nu_\lambda}| x_{\nu_\lambda}$$

$$> q^{m_{\nu_1}} a^{-1} \sum_{1}^{n} |a_\mu| - (q^{m_{\nu_1}+1} - q^{m_{\nu_1}}) \sum_{1}^{n} |a_\mu| - q^{m_{\nu_1}-m+1} \sum_{1}^{n} |a_\mu|$$

$$= q^{m_{\nu_1}} \sum_{1}^{n} |a_\mu| \, (a^{-1} - \{q-1\} - q^{-(m-1)}).$$

Therefore $$a^{-1} - q + 1 - q^{-(m-1)} < 0,$$

i.e., $$q^{-(m-1)} > a^{-1} - q + 1,$$

$$m - 1 < -\frac{\log (a^{-1} - q + 1)}{\log q},$$

provided that $$a^{-1} - q + 1 > 0.$$

---

† In the case $k = n$ everything relating to $k+1$, $k+2$, ... has to be omitted, and similarly later.

Put $$q = 1+a^{-1}-b^{-1},$$

where b is any number exceeding a. Then it follows that

$$m-1 < \frac{\log b}{\log (1+a^{-1}-b^{-1})}.$$

Hence (26) is not regular in the set of all positive numbers, and (28) holds. This completes the proof of Theorem IV.

THEOREM V. *Let $a_1, a_2, \ldots, a_n$ be complex numbers. Suppose that no non-empty subset of these numbers has the sum zero. Then the equation*

(35) $$a_1 x_1+\ldots+a_n x_n = 0$$

is not regular in the set of all complex numbers different from zero. In particular, if

(36) $$|a_{\nu_1}+a_{\nu_2}+\ldots+a_{\nu_k}| \geqq (|a_1|+|a_2|+\ldots+|a_n|)/a > 0$$

whenever $$1 \leqq k \leqq n; \quad 1 \leqq \nu_1 < \nu_2 < \ldots < \nu_k \leqq n,$$

then the degree of regularity D of (35) in the set of all complex numbers different from zero satisfies

(37) $$D < 1+\frac{1}{\delta}-\frac{\log (1+a^{-1}-q^{1+\delta}-2\pi\delta)}{\delta \log q},$$

where δ, q are any real numbers such that

(38) $$0 < 2\pi\delta < a^{-1},$$

(39) $$1 < q^{1+\delta} < 1+a^{-1}-2\pi\delta\dagger.$$

In particular

(40) $$D < 1+6\pi a+28\pi a^2 \log (3a) \quad (\textit{for every } a),$$

(41) $$D < 8\pi a^2 \log a+o(a^2 \log a) \quad (\textit{as } a \to \infty).$$

Proof. (36) implies $a \geqq 1$. The first part of the theorem follows as in the case of Theorem IV. (40) follows from (37) by putting

$$2\pi\delta = \frac{1}{3a}, \quad q^{1+\delta} = 1+\frac{1}{3a},$$

and using (32). (41) follows from (37) if we put, for every sufficiently large a,

$$2\pi\delta = \frac{1}{2a}, \quad q = 1+\frac{1}{2a}-\frac{1}{a \log a}.$$

† In other words, δ, q satisfy $0 < \delta$, $1 < q$, and have such values as to give a real value to the right-hand side of (37).

Then, as $a \to \infty$,

$$1+\frac{1}{a}-q^{1+\delta}-2\pi\delta = 1+\frac{1}{2a}-\left(1+\frac{1}{2a}-\frac{1}{a\log a}\right)^{1+(4\pi a)^{-1}}$$

$$= 1+\frac{1}{2a}-\exp\left[\left(1+\frac{1}{4\pi a}\right)\left\{\frac{1}{2a}-\frac{1}{a\log a}+O\left(\frac{1}{a^2}\right)\right\}\right]$$

$$= 1+\frac{1}{2a}-\exp\left\{\frac{1}{2a}-\frac{1}{a\log a}+O\left(\frac{1}{a^2}\right)\right\}$$

$$= 1+\frac{1}{2a}-\left(1+\frac{1}{2a}-\frac{1}{a\log a}+O\left(\frac{1}{a^2}\right)\right) = \frac{1}{a\log a}+O\left(\frac{1}{a^2}\right).$$

Hence, by (37),

$$D < 1+4\pi a-4\pi a\,\frac{\log\{(a\log a)^{-1}+O(a^{-2})\}}{\log\{1+\frac{1}{2}a^{-1}+O(a\log a)^{-1}\}}$$

$$= 1+4\pi a+4\pi a\,\frac{\log a+o(\log a)}{\frac{1}{2}a^{-1}+o(a^{-1})} = 8\pi a^2\log a+o(a^2\log a).$$

In order to prove the theorem we have to show that (37) holds, provided that (36) is true for all k, $\nu_1, \ldots, \nu_k$. We assume that (35) is m-regular in the set of all complex numbers different from zero. Choose real numbers q, δ such that $q > 1$, $\delta > 0$. Then every complex $x \neq 0$ has a unique representation

$$x = q^r e^{2\pi i t} \tag{42}$$

where $r = r(x)$, $t = t(x)$ are real numbers satisfying

$$0 \leqslant r-t < 1. \tag{43}$$

For (42) determines r uniquely and t uniquely mod 1, and (43) fixes t. Put

$$\Delta(x) = \Delta^{(m}\left(\left[\frac{t}{\delta}\right]\right) \quad (x \neq 0).$$

Then, according to the choice of m, there are numbers $x_\nu \neq 0$ for which

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta}, \quad a_1x_1+\ldots+a_nx_n = 0.$$

Let
$$r(x_\nu) = r_\nu; \quad t(x_\nu) = t_\nu; \quad \left[\frac{t_\nu}{\delta}\right] = m_\nu.$$

Then
$$x_\nu = q^{r_\nu} e^{2\pi i t_\nu},$$

$$t_\nu \leqslant r_\nu < t_\nu+1, \quad m_1 \equiv m_2 \equiv \ldots \equiv m_n \pmod{m}, \tag{44}$$

$$m_\nu\delta \leqslant t_\nu < m_\nu\delta+\delta \quad (1 \leqslant \nu \leqslant n). \tag{45}$$

Also $$m_{\nu_1} = m_{\nu_2} = \ldots = m_{\nu_k} > m_{\nu_{k+1}} \geqq \ldots \geqq m_{\nu_n},$$

where

$$\nu_1, \nu_2, \ldots, \nu_n \tag{46}$$

is a permutation of 1, 2, ..., n and k an integer, $1 \leqq k \leqq n$. Moreover, using (45) and (44), we can choose the permutation (46) in such a way that

$$m_{\nu_1}\delta \leqq t_{\nu_1} \leqq r_{\nu_1} \leqq r_{\nu_2} \leqq \ldots \leqq r_{\nu_k} < t_{\nu_k}+1 < m_{\nu_k}\delta+\delta+1 = m_{\nu_1}\delta+\delta+1.$$

Then we have, for every κ $(1 \leqq \kappa \leqq k)$,

$$\begin{aligned} |x_{\nu_\kappa}-x_{\nu_1}| &= |(q^{r_{\nu_\kappa}}e^{2\pi i t_{\nu_\kappa}}-q^{r_{\nu_1}}e^{2\pi i t_{\nu_\kappa}})+(q^{r_{\nu_1}}e^{2\pi i t_{\nu_\kappa}}-q^{r_{\nu_1}}e^{2\pi i t_{\nu_1}})| \\ &\leqq (q^{r_{\nu_\kappa}}-q^{r_{\nu_1}})+q^{r_{\nu_1}}\times 2\pi|t_{\nu_\kappa}-t_{\nu_1}| \\ &< q^{r_{\nu_1}}(q^{1+\delta}-1)+2\pi\delta q^{r_{\nu_1}} \quad (1 \leqq \kappa \leqq k). \end{aligned} \tag{47}$$

If $k < \lambda \leqq n$, then

$$m_{\nu_\lambda} \leqq m_{\nu_{k+1}} \leqq m_{\nu_k}-m = m_{\nu_1}-m,$$

and therefore

$$\begin{aligned} |x_{\nu_\lambda}| = q^{r_{\nu_\lambda}} < q^{t_{\nu_\lambda}+1} < q^{m_{\nu_\lambda}\delta+\delta+1} &\leqq q^{(m_{\nu_1}-m)\delta+\delta+1} \\ &\leqq q^{t_{\nu_1}-m\delta+\delta+1} \leqq q^{r_{\nu_1}-(m-1)\delta+1} \quad (k < \lambda \leqq n). \end{aligned} \tag{48}$$

Just as in the proof of Theorem IV, we deduce from (36), (47), (48), (38) and (39) that

$$\begin{aligned} 0 = \Big|\sum_{\mu=1}^{n} a_\mu x_\mu\Big| &= \Big|\sum_{\kappa=1}^{k} a_{\nu_\kappa}x_{\nu_1}+\sum_{\kappa=1}^{k} a_{\nu_\kappa}(x_{\nu_\kappa}-x_{\nu_1})+\sum_{\lambda=k+1}^{n} a_{\nu_\lambda}x_{\nu_\lambda}\Big| \\ &\geqq |x_{\nu_1}|\Big|\sum_{\kappa=1}^{k} a_{\nu_\kappa}\Big| - \sum_{\kappa=1}^{k}|a_{\nu_\kappa}||x_{\nu_\kappa}-x_{\nu_1}| - \sum_{\lambda=k+1}^{n}|a_{\nu_\lambda}||x_{\nu_\lambda}| \\ &> q^{r_{\nu_1}}a^{-1}\sum_{\mu=1}^{n}|a_\mu| - q^{r_{\nu_1}}(q^{1+\delta}-1+2\pi\delta)\sum_{\mu=1}^{n}|a_\mu| - q^{r_{\nu_1}-(m-1)\delta+1}\sum_{\mu=1}^{n}|a_\mu| \\ &= q^{r_{\nu_1}}\sum_{1}^{n}|a_\mu|\,(a^{-1}-q^{1+\delta}+1-2\pi\delta-q^{-(m-1)\delta+1}), \end{aligned}$$

$$q^{-(m-1)\delta+1} > a^{-1}-q^{1+\delta}+1-2\pi\delta,$$

$$(m-1)\delta-1 < -\frac{\log(a^{-1}-q^{1+\delta}+1-2\pi\delta)}{\log q},$$

$$m < 1+\frac{1}{\delta}-\frac{\log(1+a^{-1}-q^{1+\delta}-2\pi\delta)}{\delta\log q}.$$

If this holds for every m, then the admissible values of m are bounded, and the largest of them, *i.e.*, D, satisfies (37). Thus Theorem V is proved. It is obvious from the foregoing proofs that the numerical constants in (29) and (40) can be improved.

2. Consider a system of homogeneous linear equations

$$\sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m) \tag{49}$$

with arbitrary complex coefficients $a_{\mu\nu}$. Let B be a set of numbers. We say that (49) satisfies the *condition* $\Gamma(B)$ if it is possible to divide the set

$$\{1, 2, \ldots, n\}$$

into non-empty, non-overlapping subsets $S_1, S_2, \ldots, S_l$ such that, corresponding to every λ $(1 \leqslant \lambda \leqslant l)$, there exists a solution $x_\nu = \xi_\nu^{(\lambda)}$ of (49) for which

$$\xi_\nu^{(\lambda)} \prec B \quad (\nu \prec S_1, S_2, \ldots, S_\lambda),$$

$$\xi_\nu^{(\lambda)} = 0 \quad (\nu \prec S_{\lambda+1}, S_{\lambda+2}, \ldots, S_l),$$

and, moreover, all numbers $\xi_\nu^{(\lambda)}$ for ν in S_λ have the same value $\xi^{(\lambda)} \neq 0$. In other words, if we number the variables suitably we can find a matrix of the type (in the case $l = 4$)

$$\begin{pmatrix}
\xi^{(1)}, \xi^{(1)}, \ldots, \xi^{(1)}, 0, \ldots \ldots \ldots \ldots \ldots \ldots \ldots, 0 \\
\xi_1^{(2)}, \xi_2^{(2)}, \ldots, \xi_\alpha^{(2)}, \xi^{(2)}, \ldots, \xi^{(2)}, 0, \ldots \ldots \ldots \ldots, 0 \\
\xi_1^{(3)}, \xi_2^{(3)}, \ldots \ldots \ldots \ldots \xi_\beta^{(3)}, \xi^{(3)}, \ldots, \xi^{(3)}, 0, \ldots, 0 \\
\xi_1^{(4)}, \xi_2^{(4)}, \ldots \ldots \ldots \ldots \ldots \ldots \ldots \xi_\gamma^{(4)}, \xi^{(4)}, \ldots, \xi^{(4)}
\end{pmatrix}$$

in which every row constitutes a solution of (49), every $\xi_\nu^{(\lambda)}$ and $\xi^{(\lambda)}$ belongs to B, and $\xi^{(\lambda)} \neq 0$. We have

$$0 < \alpha < \beta < \gamma < n.$$

For instance, the system of equations

$$x_2 - x_1 = x_3 - x_2 = \ldots = x_{n-1} - x_{n-2} = x_n$$

satisfies $\Gamma(B)$, where B is the set of all integers. For we may put

$$l = 2; \quad S_1 = \{1, 2, \ldots, n-1\}; \quad S_2 = \{n\}.$$

The $\xi_\nu^{(\lambda)}$ are

$$\begin{pmatrix} 1, & 1, & 1, & \ldots, & 1, & 0 \\ 1, & 2, & 3, & \ldots, & n-1, & 1 \end{pmatrix}.$$

If (49) satisfies the condition $\Gamma(B)$ for some set B, then (49) satisfies $\Gamma(R_0)$, where R_0 is the ring generated by all coefficients $a_{\mu\nu}$†. This follows at once from well-known properties of systems of linear equations. Subsequently (pp. 149 *et seq.*), the structure of systems satisfying $\Gamma(B)$ will be further elucidated.

Let K be a number field, and denote by $K-\{0\}$ the set of all numbers of K except zero.

THEOREM VI. *If* (49) *is regular in* $K-\{0\}$ *then* (49) *satisfies* $\Gamma(K)$.

Before we prove this theorem we establish a simple lemma which is geometrically obvious.

LEMMA 3. *Let*

$$L_1(t),\ L_2(t),\ \ldots,\ L_r(t),\ M_1(t),\ \ldots,\ M_s(t)$$

be $r+s$ *linear forms in* $(t)=(t_1, t_2, \ldots, t_N)$. *Let* $r \geqq 0$; $s > 0$. *Suppose that*

$$L_\rho(t') = 0 \quad (1 \leqslant \rho \leqslant r) \tag{50}$$

implies that, for at least one $\sigma_0 = \sigma_0(t')$,

$$M_{\sigma_0}(t') = 0.$$

Then at least one of the forms $M_\sigma(t)$ *is a linear combination of the forms* $L_\rho(t)$.

Proof of Lemma 3. We may assume that no proper subset of the system $M_1(t), \ldots, M_s(t)$ has the same property as the whole system of forms $M_\sigma(t)$. Then $s = 1$. For if $s > 1$, then, corresponding to every σ $(1 \leqslant \sigma \leqslant s)$, there is a vector (t^σ) for which

$$L_\rho(t^{(\sigma)}) = 0 \ \ (1 \leqslant \rho \leqslant n), \qquad M_\sigma(t^{(\sigma)}) = 1, \qquad M_{\sigma'}(t^{(\sigma)}) = 0 \ \ (\sigma' \neq \sigma).$$

Then the vector $$(t') = \sum_{\sigma=1}^{s} (t^{(\sigma)})$$

satisfies (50). But, on the other hand,

$$M_\sigma(t') = M_\sigma(t^{(\sigma)}) = 1 \quad (1 \leqslant \sigma \leqslant s),$$

† We exclude the trivial case where all $a_{\mu\nu}$ vanish.

which contradicts our hypothesis. Therefore $s = 1$. In this case the lemma is an immediate consequence of well-known facts concerning linear forms.

For a later application we add the remark that (50) is required only for those $(t') = (t_1', t_2', \ldots, t_N')$ whose components t_ν' belong to the field generated by the coefficients of all forms L_ρ, M_σ.

Proof of Theorem VI. We suppose that (49) is regular in K. Then every linear combination of the equations (49) is regular in the set of all complex numbers different from zero. Therefore in every such combination certain coefficients have the sum zero (Theorem V, p. 138). If we exploit this fact for suitable linear combinations of (49), we obtain the condition $\Gamma(K)$. But before starting along these lines we replace (49) by a system of equations whose coefficients belong to K.

Let $$x_\nu = x_\nu^{(\alpha)} \quad (1 \leqslant \nu \leqslant n;\ 1 \leqslant \alpha \leqslant g)$$

be a system of a maximal number of linearly independent solutions of (49) which belong to K. Such solutions exist because (49) is regular in $K - \{0\}$. Now let

$$\sum_{\nu=0}^{n} b_{\mu\nu} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m') \tag{51}$$

be a system of linear equations whose general solution is an arbitrary linear combination of the vectors $(x_\nu^{(\alpha)})$ $(1 \leqslant \alpha \leqslant g)$. Then, as far as solutions in K are concerned, (49) and (51) have exactly the same solutions. In particular, (51) is regular in $K - \{0\}$ and therefore, *a fortiori*, regular in the set of all complex numbers different from zero. Since the $x_\nu^{(\alpha)}$ belong to K, it is possible to choose (51) so that the $b_{\mu\nu}$ belong to K.

Choose m' parameters t_μ and consider the equation

$$\sum_{\mu=1}^{m'} t_\mu \sum_{\nu=1}^{n} b_{\mu\nu} x_\nu = 0,$$

i.e.,

$$\sum_{\nu=1}^{n} R_\nu(t) x_\nu = 0, \tag{52}$$

where

$$R_\nu(t) = \sum_{\mu=1}^{m'} b_{\mu\nu} t_\mu \quad (1 \leqslant \nu \leqslant n). \tag{53}$$

Let $$M_1(t), M_2(t), \ldots, M_s(t) \quad (s = 2^n - 1)$$

be all the linear forms which are sums of forms $R_\nu(t)$ corresponding to any choice of distinct ν. For any values of the t_μ, (52) is regular in the set of all complex numbers different from zero. Hence, by Theorem V, at

least one of the numbers $M_\sigma(t)$ vanishes. Therefore, by Lemma 3 above (p. 142), with $r = 0$; $s = 2^n - 1$, at least one of the forms $M_\sigma(t)$ vanishes identically in (t). If the x_ν are suitably numbered, we may assume that, identically in (t),

$$\sum_{\nu=1}^{\alpha} R_\nu(t) = 0. \tag{54}$$

Here α is an integer and $1 \leqslant \alpha \leqslant n$.

If $\alpha = n$, we stop at this stage. If $\alpha < n$, we let the parameters t_μ be subject to the conditions

$$R_\nu(t) = 0 \quad (1 \leqslant \nu \leqslant \alpha). \tag{55}$$

Then (52) becomes

$$\sum_{\nu=\alpha+1}^{n} R_\nu(t)\, x_\nu = 0.$$

Suppose that

$$M_1'(t),\ M_2'(t),\ \ldots,\ M_{s'}'(t) \quad (s' = 2^{n-\alpha} - 1)$$

are all the sums of any number of forms $R_\nu(t)$ corresponding to distinct values of ν with $\alpha < \nu \leqslant n$. Again Theorem V shows that, for every choice of (t) satisfying (55), at least one of the numbers $M_\sigma'(t)$ vanishes, and this implies, by Lemma 3 $(r = \alpha;\ s = s')$, that at least one of the forms $M_\sigma'(t)$ $(1 \leqslant \sigma \leqslant s')$ is a linear combination of the $R_\nu(t)$ $(1 \leqslant \nu \leqslant \alpha)$. We can number the variables $x_{\alpha+1}, x_{\alpha+2}, \ldots, x_n$ so that this linear relation becomes

$$\sum_{\nu=1}^{\alpha} f_\nu R_\nu(t) + \sum_{\nu=\alpha+1}^{\beta} R_\nu(t) = 0, \tag{56}$$

identically in (t). Here β is some number satisfying $\alpha < \beta \leqslant n$, and the f_ν are constants.

If we proceed in this way, treating β as we treated α, we find (in the case $\beta < n$) a relation

$$\sum_{\nu=1}^{\beta} g_\nu R_\nu(t) + \sum_{\nu=\beta+1}^{\gamma} R_\nu(t) = 0, \tag{57}$$

where $\alpha < \beta < \gamma \leqslant n$, etc. The process stops, say, with

$$\sum_{\nu=1}^{\delta} j_\nu R_\nu(t) + \sum_{\nu=\delta+1}^{n} R_\nu(t) = 0. \tag{58}$$

We have $$1 \leqslant \alpha < \beta < \gamma < \ldots < \delta < n.$$

Since the coefficients of the R_ν belong to K, it is possible to choose the relations (56), (57), ..., (58) so that their coefficients $f_\nu, g_\nu, \ldots, j_\nu$ belong to K.

In view of (53), the relations (54), (56), (57), ..., (58) are equivalent to

$$\sum_{\nu=1}^{\alpha} b_{\mu\nu}\times 1 = 0,$$

$$\sum_{\nu=1}^{\alpha} b_{\mu\nu}f_\nu + \sum_{\nu=\alpha+1}^{\beta} b_{\mu\nu}\times 1 = 0,$$

$$\sum_{\nu=1}^{\beta} b_{\mu\nu}g_\nu + \sum_{\nu=\beta+1}^{\gamma} b_{\mu\nu}\times 1 = 0,$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\sum_{\nu=1}^{\delta} b_{\mu\nu}j_\nu + \sum_{\nu=\delta+1}^{n} b_{\mu\nu}\times 1 = 0,$$

where μ takes all values 1, 2, ..., m'. This means that the rows of the matrix

$$\begin{pmatrix} 1, & 1, & \ldots, & 1, & 0, & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ f_1, & f_2, & \ldots, & f_\alpha, & 1, & \ldots, & 1, & 0, & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ g_1, & g_2, & \ldots & \ldots & \ldots & \ldots, & g_\beta, & 1, & \ldots, & 1, & 0, & \ldots & \ldots & 0 \\ \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & \ldots \\ j_1, & j_2, & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots, & j_\delta, & 1, & \ldots, & 1 \end{pmatrix}$$

are solutions of (51). The elements of this matrix belong to K. Therefore, in view of the connection between (51) and (49), its rows are, at the same time, solutions of (49), and this amounts to saying that (49) satisfies $\Gamma(K)$. The subsets $S_1, S_2, \ldots, S_l$ of p. 141 are

$$\{1, 2, \ldots, \alpha\}, \quad \{\alpha+1, \alpha+2, \ldots, \beta\}, \quad \{\beta+1, \ldots, \gamma\}, \quad \ldots, \quad \{\delta+1, \ldots, n\}.$$

Thus Theorem VI is proved.

3. Let R be a ring of numbers, and let $R-\{0\}$ denote the set of all numbers of R except zero. The main result of this paragraph is the following theorem.

THEOREM VII. *A system of equations*

$$\sum_{\nu=1}^{n} a_{\mu\nu}x_\nu = 0 \quad (1 \leqslant \mu \leqslant m) \tag{59}$$

is regular in $R-\{0\}$ *if, and only if,* (59) *satisfies* $\Gamma(R)$†. *In particular,*

† The condition $\Gamma(R)$ was defined on p. 141.

the equation

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = 0,$$

where $a_\nu \prec R - \{0\}$, *is regular in* $R - \{0\}$ *if, and only if, some of the numbers* $a_1, a_2, \ldots, a_n$ *have the sum zero.*

Proof. Let K be the field generated by the numbers of R. To begin with the easier half of the theorem, let us assume that (59) is regular in $R - \{0\}$. Then, *a fortiori*, (59) is regular in $K - \{0\}$. Therefore, by Theorem VI, (59) satisfies $\Gamma(K)$. Let $\xi_\nu^{(\lambda)}$ be the l solutions of (59) mentioned in the definition of $\Gamma(K)$. Then

$$\xi_\nu^{(\lambda)} = \eta_\nu^{(\lambda)} / \zeta_\nu^{(\lambda)},$$

where $$\eta_\nu^{(\lambda)} \prec R; \quad \zeta_\nu^{(\lambda)} \prec R - \{0\} \quad (1 \leqslant \nu \leqslant n; \ 1 \leqslant \lambda \leqslant l).$$

Let $$N = \prod_{1 \leqslant \nu \leqslant n, \ 1 \leqslant \lambda \leqslant l} \zeta_\nu^{(\lambda)}; \quad \bar{\xi}_\nu^{(\lambda)} = N \xi_\nu^{(\lambda)}.$$

Then the numbers $\bar{\xi}_\nu^{(\lambda)}$ belong to R, and it is obvious that (59) satisfies $\Gamma(R)$.

Now let us assume that (59) satisfies $\Gamma(R)$. We have to show that (59) is regular in $R - \{0\}$. We exclude the trivial case where all $a_{\mu\nu}$ vanish. In accordance with the definition of $\Gamma(R)$ there are numbers $\xi_\nu^{(\lambda)}$ $(1 \leqslant \nu \leqslant n; \ 1 \leqslant \lambda \leqslant l)$ of R which have the properties stated on p. 141. Let B be the system of these numbers $\xi_\nu^{(\lambda)}$, and denote by R', K' the ring and the field respectively generated by the numbers of B. We shall show that (59) is regular in $K' - \{0\}$. If this fact is established, then, by Lemma 1 on p. 129, regularity in $R' - \{0\}$ will follow, and since R' is contained in R, regularity in $R - \{0\}$.

Suppose, first of all, that the number l occurring in the definition of $\Gamma(R)$ has the value 1. Then

$$\sum_{\nu=1}^{n} a_{\mu\nu} \times 1 = 0 \quad (1 \leqslant \mu \leqslant m),$$

and therefore (59) is regular, in fact, absolutely regular, in $K' - \{0\}$. Now suppose that $l = l' > 1$. We may assume that it has already been proved that all systems (59) satisfying $\Gamma(R)$ with a value $l = l' - 1$ are regular in the corresponding $K' - \{0\}$. We may suppose without loss of generality that

$$S_{l'} = \{n' + 1, n' + 2, \ldots, n\},$$

where $1 \leqslant n' < n$. Then, clearly, the system

$$\sum_{\nu=1}^{n'} a_{\mu\nu} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m) \tag{60}$$

satisfies $\Gamma(R)$ with $l = l'-1$. For we can use the same numbers $\xi_\nu^{(\lambda)}$ as for (59) but restrict ν, λ to the ranges $1 \leqslant \nu \leqslant n'$, $1 \leqslant \lambda \leqslant l'-1$. Therefore (60) is regular in $K'-\{0\}$†. Now let k be the least positive integer, if there is one, such that (59) is not k-regular in $K'-\{0\}$. We have to deduce a contradiction.

According to $\Gamma(R)$ we have

$$\sum_{\nu=1}^{n'} a_{\mu\nu}\,\xi_\nu^{(l')} + \sum_{\nu=n'+1}^{n} a_{\mu\nu}\,\xi^{(l')} = 0 \quad (1 \leqslant \mu \leqslant m), \tag{61}$$

where $\xi^{(l')} \prec R'-\{0\}$. By Theorem I (p. 128) there are finite subsets M, M' of $K'-\{0\}$ such that (59) is $(k-1)$-regular in M and (60) is k-regular in M'‡. Since (59) is not k-regular in $K'-\{0\}$, there is a distribution Δ of $K'-\{0\}$ into k classes which has the property that (59) has no solution x_ν with

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta}.$$

Let M'' be the set consisting of all numbers

$$\xi_\nu^{(l')} t y^{-1} \quad (1 \leqslant \nu \leqslant n';\ t \prec M;\ y \prec M')$$

and of the number 1. Define $\Delta'(x)$ in K' as follows:

$$\Delta'(x) = \prod_{y \prec M'} \Delta(xy) \quad (x \prec K'-\{0\}), \tag{62}$$

$$0 \not\equiv x \pmod{\Delta'} \quad (x \prec K'-\{0\}). \tag{63}$$

By Theorem III there exists a number a of K' and a positive integer d for which

$$a+dz \equiv a \pmod{\Delta'} \quad (z \prec M'').$$

In view of a remark made above (p. 135) we may postulate that the numbers a and $a+dz$ $(z \prec M'')$ differ from zero or else are all equal to zero. The second possibility is ruled out, since we included the number 1 in M''. Therefore, by (62),

$$(a+dz)\,y \equiv ay \pmod{\Delta} \quad (z \prec M'';\ y \prec M'). \tag{64}$$

Put

$$\Delta''(y) = \Delta(ay) \quad (y \prec M'). \tag{65}$$

† Strictly speaking, K' should be replaced by the field K'' belonging to (60). But K'' is contained in K', and therefore regularity in $K'-\{0\}$ holds *a fortiori*.

‡ In the case $k = 1$ statements about $k-1$ have to be omitted.

By definition of M' there are numbers y_ν of M' such that

$$y_1 \equiv y_2 \equiv \ldots \equiv y_{n'} \pmod{\Delta''}, \tag{66}$$

$$\sum_{\nu=1}^{n'} a_{\mu\nu} y_\nu = 0 \quad (1 \leqslant \mu \leqslant m); \tag{67}$$

(64), (66) and (65) imply that

$$(a+dz)\, y_\nu \equiv a y_\nu \equiv a y_1 \pmod{\Delta} \quad (1 \leqslant \nu \leqslant n';\ z \prec M''). \tag{68}$$

Also, from (67) and (61),

$$\sum_{\nu=1}^{n'} a_{\mu\nu}(a y_\nu + d\xi_\nu^{(l')} t) + \sum_{\nu=n'+1}^{n} a_{\mu\nu} \xi^{(l')} dt = 0 \quad (1 \leqslant \mu \leqslant m;\ t \prec M). \tag{69}$$

Now $$a y_\nu + d\xi_\nu^{(l')} t = (a + d\xi_\nu^{(l')} t y_\nu^{-1})\, y_\nu,$$

where $$\xi_\nu^{(l')} t y_\nu^{-1} \prec M'' \quad (1 \leqslant \nu \leqslant n';\ t \prec M).$$

Therefore, by (68),

$$a y_\nu + d\xi_\nu^{(l')} t \equiv a y_1 \pmod{\Delta} \quad (1 \leqslant \nu \leqslant n';\ t \prec M). \tag{70}$$

We now show that this implies that

$$d\xi^{(l')} t \not\equiv a y_1 \pmod{\Delta} \quad (t \prec M). \tag{71}$$

If (71) is not true, then, for some number t_0 of M,

$$d\xi^{(l')} t_0 \equiv a y_1 \pmod{\Delta}. \tag{72}$$

But then (69), (70) (with $t = t_0$) and (72) show that, contrary to our initial assumption about Δ, the system (59) has a solution whose numbers belong to the same class of Δ. Therefore (71) is established.

If $k = 1$, then (71) is plainly impossible; for then all numbers belong to the same class of Δ. When $k > 1$ put

$$\Delta'''(x) = \Delta(d\xi^{(l')} x) \quad (x \prec M).$$

Then, by (71), $$|\Delta'''| \leqslant k-1.$$

Hence, from the definition of M, there are numbers x_ν of M satisfying (59) and

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta'''}.$$

Then

$$\sum_{\nu=1}^{n} a_{\mu\nu} d\xi^{(l')} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m),$$

$$d\xi^{(l')} x_\nu \equiv d\xi^{(l')} x_1 \pmod{\Delta} \quad (1 \leqslant \nu \leqslant n)$$

and this, again, contradicts the definition of Δ. This completes the proof of Theorem VII.

4. We note a corollary of Theorem VII.

A system $$\sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m)$$

is regular in the set of all complex numbers other than 0 *if, and only if, it is regularin* $R_0 - \{0\}$, *where* R_0 *is the ring generated by all coefficients* $a_{\mu\nu}$.

(*We have to exclude the trivial case where all* $a_{\mu\nu} = 0$.)

It is, however, not true that, corresponding to every system of equations which is regular in some set, there exists a smallest ring R^* such that it is regular in $R^* - \{0\}$. For consider the equation ($m = 1$; $n = 4$)

$$\alpha x_1 + \beta x_2 - (\alpha+\beta) x_3 - \alpha\beta x_4 = 0, \tag{73}$$

where $\alpha = \sqrt{2}$; $\beta = \sqrt{3}$. The only possible sets S_1, S_2, ..., consistent with the definition of Γ (p. 141) are

$$S_1 = \{1, 2, 3\}; \quad S_2 = \{4\}.$$

The existence of the solutions

$$(1, 1, 1, 0); \quad (0, \alpha, 0, 1); \quad (\beta, 0, 0, 1)$$

of (73) shows that (73) is regular in $R_\alpha - \{0\}$ as well as in $R_\beta - \{0\}$, where R_α, R_β are the rings generated by α, 1 and β, 1 respectively. The common numbers of R_α and R_β form the ring R_0 of all rational integers. But (73) is not regular in $R_0 - \{0\}$. For this would imply that $\Gamma(R_0)$ holds, *i.e.* that there are rational integers x_1, x_2, x_3, x_4, with $x_4 \neq 0$, such that

$$\alpha x_1 + \beta x_2 - (\alpha+\beta) x_3 - \alpha\beta x_4 = 0,$$

$$(x_1 - x_3)\sqrt{2} + (x_2 - x_3)\sqrt{3} = x_4\sqrt{6},$$

$$2(x_1 - x_3)(x_2 - x_3)\sqrt{6} = 6x_4^2 - 2(x_1 - x_3)^2 - 3(x_2 - x_4)^2,$$

which is easily seen to be impossible.

5. Let R be a ring of numbers (not consisting of zero only), and K be the field generated by R. We want to bring out more clearly the significance of the condition $\Gamma(R)$. From considerations at the beginning of the proof of Theorem VII, it is obvious that $\Gamma(R)$ and $\Gamma(K)$ are equivalent. We can easily write down a general class of systems of linear equations which satisfy $\Gamma(K)$, viz.

$$\sum_{\nu=1}^{n-m} c_{\mu\nu} x_\nu - x_{n-m+\mu} = 0 \quad (1 \leqslant \mu \leqslant m). \tag{74}$$

Here $1 \leqslant m < n$, and the $c_{\mu\nu}$ are arbitrary numbers of K subject to the two conditions:

(i) no row

$$c_{\mu 1}, c_{\mu 2}, \ldots, c_{\mu, n-m} \quad (1 \leqslant \mu \leqslant m) \tag{75}$$

contains zeros only;

(ii) the first non-zero number of every row (75) has the value 1.

Let us call systems (74) of this kind *T-systems.*

In order to prove that (74) satisfies $\Gamma(K)$, we may assume (74) to be of the form

$$\left\{\begin{aligned}
&x_{s_1} + \sum_{\nu=s_1+1}^{n-m} c_{\mu\nu} x_\nu - x_{n-m+\mu} = 0 \quad (1 \leqslant \mu \leqslant a_1),\\
&x_{s_2} + \sum_{\nu=s_2+1}^{n-m} c_{\mu\nu} x_\nu - x_{n-m+\mu} = 0 \quad (a_1 < \mu \leqslant a_2),\\
&\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots\\
&x_{s_{l-1}} + \sum_{\nu=s_{l-1}+1}^{n-m} c_{\mu\nu} x_\nu - x_{n-m+\mu} = 0 \quad (a_{l-2} < \mu \leqslant a_{l-1}),
\end{aligned}\right. \tag{76}$$

where $$l \geqslant 2; \quad 1 \leqslant s_1 < s_2 < \ldots < s_{l-1} \leqslant n-m;$$

$$1 \leqslant a_1 < a_2 < \ldots < a_{l-1} = m.$$

Then we put

$$S_1 = \{s_1, n-m+1, n-m+2, \ldots, n-m+a_1\},$$

$$S_2 = \{s_2, n-m+a_1+1, n-m+a_1+2, \ldots, n-m+a_2\},$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$S_{l-1} = \{s_{l-1}, n-m+a_{l-2}+1, n-m+a_{l-2}+2, \ldots, n-m+a_{l-1}\},$$

while the last set S_l consists of all indices out of $1, 2, \ldots, n$ which do not occur in any of the sets $S_1, S_2, \ldots, S_{l-1}$. A moment's consideration shows that we can choose numbers $\xi_\nu^{(\lambda)}$ which are consistent with the definition of $\Gamma(K)$.

In a certain sense (76) is the most general system which satisfies $\Gamma(K)$. For we have the following result.

Corresponding to every system

$$\sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = 0 \quad (1 \leqslant \mu \leqslant m') \tag{77}$$

which satisfies $\Gamma(K)$, *there is a* T*-system* (74) *such that every solution of* (74) *is at the same time a solution of* (77).

In other words, by adding suitable equations to those of (77), which is supposed to satisfy $\Gamma(K)$, we can obtain a T-system (or a system equivalent to a T-system).

In proving this statement we may assume, without loss of generality, that the sets S_λ corresponding to (77) are

$$S_1 = \{1, 2, \ldots, \beta_1\},$$

$$S_2 = \{\beta_1+1, \beta_1+2, \ldots, \beta_2\},$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$S_l = \{\beta_{l-1}+1, \beta_{l-1}+2, \ldots, \beta_l\},$$

where $$l \geqslant 1; \quad 1 \leqslant \beta_1 < \beta_2 < \ldots < \beta_l = n.$$

Consequently there is a matrix

$$(78) \qquad \begin{pmatrix} 1, & 1, & \ldots, & 1, & 0, & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots, & 0 \\ \xi_1^{(2)}, & \xi_2^{(2)}, & \ldots, & \xi_{\beta_1}^{(2)}, & 1, & \ldots, & 1, & 0, & \ldots & \ldots & \ldots & \ldots & \ldots, & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & & & \\ \xi_1^{(l)}, & \xi_2^{(l)}, & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \xi_{\beta_{l-1}}^{(l)}, & 1, & \ldots, & 1 & \end{pmatrix}$$

every row of which is a solution of (77). The $\xi_\nu^{(\lambda)}$ belong to K. In (78) the columns with indices $\beta_1, \beta_2, \ldots, \beta_l$ are linearly independent, while every one of the remaining columns is a linear combination of these l columns. Moreover, these linear relations are expressible in the form

$$(79) \qquad \begin{cases} x_\mu = x_{\beta_1} + \sum_{\lambda=2}^{l} c_{\mu\lambda} x_{\beta_\lambda} & (1 \leqslant \mu < \beta_1), \\ x_\mu = x_{\beta_2} + \sum_{\lambda=3}^{l} c_{\mu\lambda} x_{\beta_\lambda} & (\beta_1 < \mu < \beta_2), \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots & \\ x_\mu = x_{\beta_l} & (\beta_{l-1} < \mu < \beta_l), \end{cases}$$

where $c_{\mu\lambda} \prec K$. (79) is a complete set of linearly independent relations between the columns of (78). Hence every vector $(x_1, x_2, \ldots, x_n)$ which satisfies (79) is a linear combination of the rows of (78) and is therefore a solution of (77). But (79) is a T-system, as is seen by ordering the variables as follows:

$$x_{\beta_1}, x_{\beta_2}, \ldots, x_{\beta_l}, x_1, x_2, \ldots, x_{\beta_1-1}, x_{\beta_1+1}, x_{\beta_1+2}, \ldots, x_{\beta_2-1}, x_{\beta_2+1}, \ldots, x_{\beta_l-1}.$$

This proves our statement about the connection between (77) and T-systems.

Questions about the regularity of a system of homogeneous linear equations in the set of all positive rational integers or in the set of all complex numbers $re^{i\phi}$, where $r>0$ and ϕ belongs to a given interval $\phi_0<\phi<\phi_1$, and many more, are settled by the following theorem.

THEOREM VIII. *Let R be a ring of complex numbers, and let A be a subset of $R-\{0\}$ which has the following property. There is a finite number of elements $d_1, d_2, \ldots, d_k$ of $R-\{0\}$ such that every x of $R-\{0\}$ is expressible as*

$$x = ad_\kappa, \tag{80}$$

where $a \prec A$; $1 \leqslant \kappa \leqslant k$. Then regularity of

$$\sum_{\nu=1}^{n} a_{\mu\nu}x_\nu = 0 \quad (1 \leqslant \mu \leqslant m) \tag{81}$$

in A is equivalent to regularity of (81) *in $R-\{0\}$.*

For instance, in the case of all numbers $re^{i\phi}$, where $r>0$; $\phi_0<\phi<\phi_1$, we choose an integer $k>2\pi/(\phi_1-\phi_0)$ and put

$$d_\kappa = e^{2\pi i\kappa/k} \quad (1 \leqslant \kappa \leqslant k).$$

Of course, the representation (80) need not be unique.

Proof. Suppose that (81) is regular in $R-\{0\}$. We have to show that (81) is regular even in A. Let $\Delta(x)$ be defined in A. Given any x of $R-\{0\}$, we define $\lambda=\lambda(x)$ as being the least index κ for which xd_κ^{-1} belongs to A: λ exists. Put

$$\Delta'(x) = \Delta^{(0)}\big(\lambda(x)\big)\,\Delta(xd_{\lambda(x)}^{-1}) \quad (x \prec R-\{0\}).$$

Then there are numbers x_ν satisfying (81) and

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta'}. \tag{82}$$

(82) implies that

$$\lambda(x_1)=\lambda(x_2)=\ldots=\lambda(x_n)=\lambda_0,$$

say, and

$$x_1 d_{\lambda_0}^{-1} \equiv x_2 d_{\lambda_0}^{-1} \equiv \ldots \equiv x_n d_{\lambda_0}^{-1} \pmod{\Delta};$$

while (81) yields

$$\sum_\nu a_{\mu\nu}x_\nu d_{\lambda_0}^{-1} = 0 \quad (1 \leqslant \mu \leqslant m).$$

This proves the theorem.

3. *Non-homogeneous linear equations.*

1. Let A_0 be the field of all algebraic numbers and C_0 that of all complex numbers.

THEOREM IX. *A system*

$$\sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = b_\mu \quad (1 \leqslant \mu \leqslant m) \tag{83}$$

is regular in A_0 if, and only if, it is absolutely regular in A_0.

THEOREM X. *A system* (83) *with algebraic coefficients $a_{\mu\nu}$ (and with the b_μ arbitrary complex numbers) is regular in C_0 if, and only if, it is absolutely regular in C_0*†).

In proving the last two theorems it is sufficient to show that regularity implies absolute regularity in the sets in question.

Theorem IX follows from Theorem X. For let us assume that (83) is regular in A_0. Then there are algebraic numbers $x_\nu^{(0)}$ satisfying

$$\sum_\nu a_{\mu\nu} x_\nu^{(0)} = b_\mu \quad (1 \leqslant \mu \leqslant m).$$

This simply means that (83) is 1-regular in A_0. Therefore (83) is equivalent to

$$\sum_\nu a_{\mu\nu}(x_\nu - x_\nu^{(0)}) = 0 \quad (1 \leqslant \mu \leqslant m).$$

If we use the argument which leads from (49) to (51) we find a system

$$\sum_{\nu=1}^{n} a'_{\mu\nu} y_\nu = 0 \quad (1 \leqslant \mu \leqslant m') \tag{84}$$

with algebraic coefficients $a'_{\mu\nu}$ such that (84) and

$$\sum_{\nu=1}^{n} a_{\mu\nu} y_\nu = 0 \quad (1 \leqslant \mu \leqslant m)$$

have exactly the same solutions in A_0. Therefore

$$\sum_\nu a'_{\mu\nu}(x_\nu - x_\nu^{(0)}) = 0 \quad (1 \leqslant \mu \leqslant m') \tag{85}$$

and (83) have the same solutions in A_0. In particular, (85) is regular in A_0 and *a fortiori* regular in C_0. Now apply Theorem X to (85). It follows that (85) is absolutely regular in C_0. Since $a'_{\mu\nu}$, $x_\nu^{(0)}$ are numbers of A_0 we conclude further that (85) is absolutely regular even in A_0, and finally that (83) is absolutely regular in A_0.

† Absolute regularity was defined on p. 127.

It remains to prove Theorem X. We note that absolute regularity in a set M of a single equation

$$a_1x_1+a_2x_2+\ldots+a_nx_n=b$$

means that either

$$a_1+a_2+\ldots+a_n=0$$

or

$$s=a_1+\ldots+a_n\neq 0;\quad b/s\prec M.$$

LEMMA 4. *Suppose that* (83) *has the following property. For every choice of numbers* t_μ *belonging to the field* K *generated be the* $a_{\mu\nu}$, *the equation*

$$\sum_{\mu=1}^{m} t_\mu \sum_{\nu=1}^{n} a_{\mu\nu}x_\nu = \sum_{\mu=1}^{m} t_\mu b_\mu$$

is absolutely regular in M. *Then* (83) *is absolutely regular in* M.

To prove Lemma 4, put $s_\mu=\sum_\nu a_{\mu\nu}$. We may assume that

$$s_\mu\neq 0\ \ (1\leqslant\mu\leqslant m'),\qquad s_\mu=0\ \ (m'<\mu\leqslant m),$$

where $0\leqslant m'\leqslant m$. By hypothesis, any numbers t_μ of K for which

$$\sum_\mu s_\mu t_\mu=0$$

satisfy

$$\sum_\mu b_\mu t_\mu=0.$$

Therefore, by Lemma 3 on p. 142 ($r=s=1$: see the remark at the end of the proof of Lemma 3),

$$x'\sum_\mu s_\mu t_\mu=\sum_\mu b_\mu t_\mu,$$

identically in the t_μ. Here x' is some constant. Hence

$$\sum_\nu a_{\mu\nu}x'=b_\mu\quad(1\leqslant\mu\leqslant m).$$

If $m'>0$ then $x'=b_1/s_1\prec M$.

For the equation $\sum_\nu a_{1\nu}x_\nu=b_1$

is absolutely regular in M. If $m'=0$, *i.e. if*

$$s_\mu=0\quad(1\leqslant\mu\leqslant m),$$

then

$$b_\mu=0\quad(1\leqslant\mu\leqslant m);$$

and therefore

$$\sum_\nu a_{\mu\nu}x''=b_\mu\quad(1\leqslant\mu\leqslant m),$$

where x'' is any arbitrary number of M. Thus the lemma is proved.

In view of Lemma 4, it is sufficient to consider the case $m = 1$ of Theorem X. Let us therefore assume that $a_1, a_2, \ldots, a_n$ belong to A_0, that b is an arbitrary number and that

$$a_1+a_2+\ldots+a_n = 0; \quad b \neq 0. \tag{86}$$

We have to show that

$$a_1 x_1+\ldots+a_n x_n = b \tag{87}$$

is not regular in C_0. We may assume that no a_ν vanishes.

By means of transfinite induction we can find a "basis" $\xi_1, \xi_2, \ldots, \xi_\alpha, \ldots$ of all complex numbers with respect to the field K generated by $a_1, a_2, \ldots, a_n$. This means that every complex number x is uniquely representable in the form

$$x = \sum_\alpha x_\alpha' \xi_\alpha, \tag{88}$$

where α runs through certain ordinal numbers and x_α' belongs to K. For every x only a finite number of "coordinates" x_α' differ from zero. Let

$$b = \sum_\alpha b_\alpha' \xi_\alpha$$

be the representation (88) in the case $x = b$. Since $b \neq 0$, at least one of the numbers b_α' differs from zero. There is no loss of generality in assuming that $b_1' \neq 0$. The numbers $a_1, a_2, \ldots, a_n, b_1'$ are algebraic. If we multiply (87) by a suitable number of K, we can obtain a case where these $n+1$ numbers are algebraic integers of K. We may therefore assume that (87) is such that these numbers are algebraic integers. Let $\omega_1, \omega_2, \ldots, \omega_p$ be a minimal basis of K, so that every algebraic integer of K has a unique representation in the form

$$r_1\omega_1+r_2\omega_2+\ldots+r_p\omega_p, \tag{89}$$

where the r_λ are rational integers. Then every number of K has a unique representation (89) with rational coefficients r_λ.

Let $\mathfrak{p}$ be a prime ideal in K which is not a divisor of b_1'. Define $\Delta(x)$, for all algebraic integers of K, by means of the rule that

$$x \equiv y \pmod{\Delta}$$

if, and only if,

$$x \equiv y \pmod{\mathfrak{p}}.$$

Then no class of Δ contains a solution of

$$a_1 x_1+\ldots+a_n x_n = b_1'. \tag{90}$$

For (90) and

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta} \tag{91}$$

would lead to the contradiction

$$b_1' = \Sigma a_\nu x_\nu = \Sigma a_\nu (x_\nu - x_1) \equiv 0 \pmod{\mathfrak{p}}.$$

Define, for every x of K, a function $f(x)$ by means of

$$f(x) = \sum_{\lambda=1}^{p} |r_\lambda|,$$

where

$$x = \sum_{\lambda=1}^{p} r_\lambda \omega_\lambda \quad (r_\lambda \text{ rational}). \tag{92}$$

Then, for any x_ν satisfying (91),

$$f(\Sigma a_\nu x_\nu - b_1') \geqq 1. \tag{93}$$

Now choose an integer N satisfying

$$N > \sum_{1 \leqq \nu \leqq n,\ 1 \leqq \lambda \leqq p} f(a_\nu \omega_\lambda); \tag{94}$$

and put, for every x of K satisfying (92),

$$\Delta'(x) = \Delta\left(\sum_\lambda [r_\lambda] \omega_\lambda\right) \prod_{\lambda=1}^{p} \Delta^{(N)}([N r_\lambda])\dagger. \tag{95}$$

We now show that no class of Δ' contains a solution of (90). Suppose that

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta'}, \tag{96}$$

$$a_1 x_1 + \ldots + a_n x_n = b_1', \tag{97}$$

$$x_\nu = \sum_\lambda r_{\nu\lambda} \omega_\lambda \quad (r_{\nu\lambda} \text{ rational}).$$

(96) and (95) imply that

$$\sum_\lambda [r_{\nu\lambda}] \omega_\lambda \equiv \sum_\lambda [r_{1\lambda}] \omega_\lambda \pmod{\Delta} \quad (1 \leqslant \nu \leqslant n), \tag{98}$$

$$[N r_{\nu\lambda}] \equiv [N r_{1\lambda}] \pmod{N} \quad (1 \leqslant \nu \leqslant n;\ 1 \leqslant \lambda \leqslant p),$$

$$\frac{1}{N}[N r_{\nu\lambda}] - [r_{\nu\lambda}] = \frac{1}{N}[N r_{\nu\lambda}] - \left[\frac{1}{N}[N r_{\nu\lambda}]\right]$$

$$= \frac{1}{N}[N r_{1\lambda}] - \left[\frac{1}{N}[N r_{1\lambda}]\right] = \frac{1}{N}[N r_{1\lambda}] - r_{1\lambda} = d_\lambda,$$

---

† $\Delta^{(N)}$ was defined on p. 124.

say, for $$1 \leqslant \nu \leqslant n, \quad 1 \leqslant \lambda \leqslant p,$$

$$\sum_\nu a_\nu \sum_\lambda [r_{\nu\lambda}]\,\omega_\lambda = \sum_\nu a_\nu \sum_\lambda \frac{1}{N}\,[Nr_{\nu\lambda}]\,\omega_\lambda - \sum_\nu a_\nu \sum_\lambda d_\lambda \omega_\lambda$$

$$= \sum_\nu a_\nu \sum_\lambda \frac{1}{N}\,[Nr_{\nu\lambda}]\,\omega_\lambda \quad \text{[by (86)]}$$

$$= \sum_\nu a_\nu \sum_\lambda r_{\nu\lambda}\omega_\lambda - \sum_\nu a_\nu \sum_\lambda r'_{\nu\lambda}\omega_\lambda,$$

where $r'_{\nu\lambda}$ is rational, $0 \leqslant r'_{\nu\lambda} < 1/N$. Hence, from (98) and (93),

$$1 \leqslant f\Big(\sum_\nu a_\nu \sum_\lambda [r_{\nu\lambda}]\,\omega_\lambda - b_1'\Big) = f\Big(\sum_\nu a_\nu \sum_\lambda r_{\nu\lambda}\omega_\lambda - \sum_\nu a_\nu \sum_\lambda r'_{\nu\lambda}\omega_\lambda - b_1'\Big)$$

$$= f\Big(-\sum_\nu a_\nu \sum_\lambda r'_{\nu\lambda}\omega_\lambda\Big) \leqslant \sum_{\nu,\lambda} |r'_{\nu\lambda}|\, f(a_\nu \omega_\lambda) \leqslant \frac{1}{N} \sum_{\nu,\lambda} f(a_\nu \omega_\lambda).$$

The last inequality contradicts (94). Therefore, as stated above, no class of Δ' contains a solution of (90).

We now proceed to define a distribution of C_0. Put, for every complex x,

$$g(x) = x_1',$$

where x_1' is the first "coordinate" in the representation (88). $g(x)$ has the following properties:

$$g(x) \prec K, \quad g(b) = b_1' \neq 0, \quad g\Big(\sum_{\nu=1}^{n} a_\nu x_\nu\Big) = \sum_\nu a_\nu g(x_\nu)$$

for all complex x_ν.

Put $$\Delta''(x) = \Delta'\big(g(x)\big) \quad (x \prec C_0).$$

Suppose that, for some x_ν of C_0,

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta''}.$$

Then $$g(x_\nu) \equiv g(x_1) \pmod{\Delta'} \quad (1 \leqslant \nu \leqslant n).$$

Hence, by the definition of Δ',

$$\sum_\nu a_\nu g(x_\nu) \neq g(b),$$

i.e., $$g\Big(\sum_\nu a_\nu x_\nu\Big) \neq g(b);$$

and therefore $$\sum_\nu a_\nu x_\nu \neq b.$$

This completes the proofs of Theorems IX and X.

2. The foregoing proof made use of transfinite induction in order to define a certain distribution of all complex numbers. In the case of rational coefficients $a_{\mu\nu}$ (the b_μ may be arbitrary complex numbers) it is possible to eliminate transfinite induction. In this section we give a proof of the assertion of Theorem X in the special case of rational coefficients $a_{\mu\nu}$ or even in the slightly more general case where

$$a_{\mu\nu} = a'_{\mu\nu} + ia''_{\mu\nu} \quad (a'_{\mu\nu},\ a''_{\mu\nu} \text{ rational}).$$

In view of Lemma 4 it is sufficient to prove the following proposition.

Let

$$a_\nu = a_\nu' + ia_\nu'' \quad (1 \leqslant \nu \leqslant n),$$

where a_ν', a_ν'' are rational integers. Suppose that

$$a_1 + \ldots + a_n = 0,$$

while b is a complex number different from zero. Then

$$a_1 x_1 + \ldots + a_n x_n = b \tag{99}$$

is not regular in C_0.

Define, for any

$$x = x' + ix'' \quad (x',\ x'' \text{ real}),$$

$$[x] = [x'] + i[x''].$$

Choose positive integers m', m'' such that

$$m' > \sqrt{2}\sum_\nu |a_\nu| + |b|; \quad m'' > \sqrt{2}\,|b|^{-1}\sum_\nu |a_\nu|.$$

Put, for every such x,

$$\Delta^*(x) = \Delta^{(m'\,m'')}([m''\,x'])\,\Delta^{(m'\,m'')}([m''\,x'']).$$

Then our proof is complete when we can show that no class of Δ^* contains a solution of (99). Suppose that

$$x_1 \equiv x_2 \equiv \ldots \equiv x_n \pmod{\Delta^*}.$$

Then (using an obvious notation)

$$[m''\,x_\nu] \equiv [m''\,x_1] \pmod{m'\,m''},$$

$$[m''\,x_\nu] - [m''\,x_1] = k_\nu\, m'\, m'' \quad (1 \leqslant \nu \leqslant n), \tag{100}$$

where

$$k_\nu = k_\nu' + ik_\nu'' \quad (k_\nu',\ k_\nu'' \text{ rational integers}).$$

(100) implies that

$$\frac{1}{m''}[m''x_\nu]-\frac{1}{m''}[m''x_1]=k_\nu m',$$

$$\left[\frac{1}{m''}[m''x_\nu]\right]-\left[\frac{1}{m''}[m''x_1]\right]=k_\nu m',$$

$$[x_\nu]-[x_1]=k_\nu m'.$$

Also, $$m''x_\nu-m''x_1=k_\nu m'm''+r_\nu,$$

where $$|r_\nu|<\sqrt{2}.$$

Case (1): Let $$\Sigma a_\nu[x_\nu]\neq 0.$$

Then

$$|\Sigma a_\nu[x_\nu]|=|\Sigma a_\nu([x_\nu]-[x_1])|=|\Sigma a_\nu k_\nu m'|\geqslant m',$$

$$|\Sigma a_\nu x_\nu|\geqslant|\Sigma a_\nu[x_\nu]|-|\Sigma a_\nu(x_\nu-[x_\nu])|\geqslant m'-\Sigma|a_\nu|\sqrt{2}>|b|.$$

Case (2): Let $$\Sigma a_\nu[x_\nu]=0.$$

Then

$$|\Sigma a_\nu x_\nu|=|\Sigma a_\nu\{(x_\nu-x_1)-([x_\nu]-[x_1])\}|$$

$$=\left|\Sigma a_\nu\left(\frac{k_\nu m'm''+r_\nu}{m''}-k_\nu m'\right)\right|=\left|\Sigma\frac{a_\nu r_\nu}{m''}\right|\leqslant\frac{1}{m''}\Sigma|a|\sqrt{2}<|b|.$$

Hence in either case

$$a_1x_1+a_2x_2+\ldots+a_nx_n\neq b.$$

4. *ω-regularity.*

A systems of conditions

(101) $$S(x_1,\ \ldots,\ x_n)=0$$

was called *ω-regular* in a set M if, given any distribution Δ of M into denumerably many classes, there is always a solution of (101) satisfying

$$x_1\equiv x_2\equiv\ldots\equiv x_n\pmod{\Delta}.$$

Of the three propositions:

(i) (101) is k-regular in C_0 for every $k=1,\ 2,\ 3,\ \ldots$;

(ii) (101) is ω-regular in C_0;

(iii) (101) is absolutely regular in C_0;

(i) is weaker than (ii), and (ii) is weaker than (iii). For the equation

$$x_1+x_2-x_3=0$$

satisfies (i) but not (ii) (as follows from Theorem XI below), and the condition

$$x_1 \neq x_2$$

satisfies (ii) but not (iii). We prove that in the case of linear equations or, more generally, of conditions whose solutions form a closed set, (ii) and (iii) are equivalent.

THEOREM XI. *A system of equations*

$$\sum_{\nu=1}^{n} a_{\mu\nu}x_\nu = b_\mu \quad (1 \leqslant \mu \leqslant m) \tag{102}$$

is ω-regular in a set M of complex numbers if, and only if, (102) *is absolutely regular in M.*

Proof. Suppose that (102) is not absolutely regular in M. We have to show that (102) is not ω-regular in M. We call a circle

$$|x-(c'+ic'')|<r \tag{103}$$

a *rational circle* if c', c'', r are rational numbers. Every point z of M is contained in a rational circle which does not contain any solution of (102). For, if this is not true for some $z=z_0$ of M, then every rational circle (103) which contains z_0 also contains a solution x_ν of (102), and making $r \to 0$ shows that

$$\sum_\nu a_{\mu\nu}z_0 = b_\mu \quad (1 \leqslant \mu \leqslant m),$$

i.e. that (102) is absolutely regular in M. The rational circles are enumerable. Let R_1, R_2, ... be a sequence containing every rational circle. Define, for every z of M, $f(z)$ as being the least λ such that z lies in R_λ but, at the same time, R_λ contains no solution of (102). Then, obviously no class of

$$\Delta^*(z)=\Delta^{(0)}\big(f(z)\big) \quad (z \prec M)\dagger$$

contains a solution of (102). Therefore (102) is not ω-regular in M.

The University,
Sheffield.

† $\Delta^{(0)}$ was defined on p. 124.

RELATIVISTIC MECHANICS: (III) MECHANICS IN 4-SPACE, AND CONSERVATIVE FIELDS OF FORCE

By A. G. WALKER.

[Received 6 January, 1941.—Read 29 May, 1941.]

1. In previous papers [1]† we discussed as generally as possible the descriptions of physical objects, particularly those occurring in dynamics. These have so far been restricted to 3-space maps, S_3 with t-time and Σ_3 with τ-time, but in the present paper the results are extended to 4-space maps S_4 and Σ_4. The reasons for this are, first, to discover whether the formulae in relativity mechanics are in any way simplified or made more significant by using 4-space instead of 3-space. Secondly, when physical objects, formulae, etc., are described in 4-space, it is then possible by tensor transformations to pass from the special coordinate systems already used to any other, such as a "private" system of the kind used by Milne.

Our procedure will be to extend definitions and formulae from Σ_3 to Σ_4, both with τ-time, and then to transform to S_4 and t-time. It is hoped that all the various systems and descriptions will not confuse the reader, but they must be considered as generally as possible in order that the simplest and most suitable scheme of mechanics may be disclosed. When this has been done, the others can be disregarded.

2. It may be useful to recall the notation and formulae given in the preceding papers. The metrics of S_3 and S_4 are respectively

$$de^2 = k_{ij}\,dz^i\,dz^j, \quad ds^2 = dt^2 - t^2\,de^2, \tag{1}$$

and the *special coordinate system* is that for which

$$k_{ij} = (1+\tfrac{1}{4}K|z|^2)^{-2}\,\delta_j^i. \tag{1'}$$

† Numerals in square brackets refer to the list of references at the end of this paper.

The metrics of Σ_3 and Σ_4 are respectively

$$d\epsilon^2 = k_{ij}\, d\zeta^i d\zeta^j, \quad d\sigma^2 = d\tau^2 - c^{-2} d\epsilon^2, \tag{2}$$

and here the special coordinate system is the system for which

$$k_{ij} = \left(1 + \frac{K}{4c^2 t_0^2} |\zeta|^2\right)^{-2} \delta_j^i. \tag{2'}$$

The two coordinate systems are connected by the relations

$$\tau = t_0 + t_0 \log \frac{t}{t_0}, \quad \zeta^i = c t_0 z^i, \tag{3}$$

and an important invariant in both systems is

$$\eta = \left(\frac{ds}{dt}\right)^2 = \left(\frac{d\sigma}{d\tau}\right)^2. \tag{4}$$

The proper mass m, the relative mass M, and the energy E are connected by the relations

$$M = m\phi(\eta), \quad E = Mc^2. \tag{5}$$

The time intervals are dT, dT' in t- and τ-mechanics respectively, where

$$dT = \frac{dt}{\theta(\eta)}, \quad dT' = \frac{t_0}{t} dT = \frac{d\tau}{\theta(\eta)}. \tag{6}$$

The momentum vector is p^i in both t- and τ-mechanics, where

$$p^i = Mct \frac{dz^i}{dT} = M \frac{d\zeta^i}{dT'}. \tag{7}$$

The absolute (tensor) differential in 3-space is the same for both S_3 and Σ_3 and is denoted by δ. The force vectors are F^i, F'^i in t- and τ-mechanics respectively, where

$$F^i = \frac{\delta p^i}{dT}, \quad F'^i = \frac{t}{t_0} F^i = \frac{\delta p^i}{dT'}. \tag{8}$$

For the special coordinates z^i in S_3 and ζ^i in Σ_3, the position vectors are P^i, P'^i where

$$P^i = ctz^i, \quad P'^i = \frac{t_0}{t} P^i = \zeta^i. \tag{9}$$

The moment tensors are m_{ij} in S_3 and m'_{ij} in Σ_3 where

$$m_{ij} = -\epsilon_{ijk} P^k, \quad m'_{ij} = \frac{t_0}{t} m_{ij} = -\epsilon_{ijk} P'^k. \tag{10}$$

The angular equations of motion are

$$\frac{1}{t} m_{ij} F^j = \frac{\delta}{dT}\left(\frac{1}{t} m_{ij} p^j\right), \quad m'_{ij} F'^j = \frac{\delta}{dT'}(m'_{ij} p^j). \tag{11}$$

The above system of mechanics includes for generality functions $\theta(\eta)$ and $\phi(\eta)$. Two particular systems are, first, the *M-system* which is used by Milne and in which

$$M\text{-system}: \qquad \theta \equiv \phi \equiv \eta^{-\frac{1}{2}}. \tag{12}$$

Here, $dT = ds$, $dT' = d\sigma$, and the relative mass is $M = m\eta^{-\frac{1}{2}}$. Secondly, there is the *W-system*, which was proposed in [1] and which appeared to be the system most similar to Newtonian mechanics. This is given by

$$W\text{-system}: \qquad \theta \equiv 1, \quad \phi \equiv \eta^{-\frac{1}{2}}, \tag{13}$$

so that $dT = dt$, $dT' = d\tau$, and $M = m\eta^{-\frac{1}{2}}$ as before. It should be remembered that these systems do not differ essentially since they are merely different ways of recording observations. Any preference for one of these systems is on grounds of appearance and simplicity.

3. When extending definitions, etc., from 3-space Σ_3 to 4-space Σ_4 we first refer to the special coordinate system, *i.e.* τ, ζ^i, and then express the results in tensor form. The results then apply at once to any system of coordinates, since it is assumed that objects are to be described by tensors of the first kind in 4-space as in 3-space (see [1], § 11). A general system of coordinates in Σ_4 is† x^α, the metric being

$$d\sigma^2 = \gamma_{\alpha\beta} dx^\alpha dx^\beta. \tag{14}$$

A method for describing objects in 4-space is not implicit in the set of rules already used in 3-space, but two obvious requirements are as follows:

(i) Physical objects must be represented by tensors of the second kind in 4-space, *i.e.* they must transform as tensors in passing from one observer to another (see [1], § 11).

(ii) The tensor object in public 3-space must be the projection of the equivalent object in 4-space‡.

† Throughout this paper Greek suffixes take the values 0, 1, 2, 3, and Roman suffixes take the values 1, 2, 3 only.

‡ The point ζ^i in Σ_3 is the projection of the point τ, ζ^i in Σ_4, whence the interval in 3-space is the projection of the interval in 4-space. The desirability of (ii) can now be seen to follow from the fact that small displacements are vectors directly associated with certain physical objects.

If† $\underset{4}{\lambda^a}$ is a vector object in Σ_4 equivalent to an object λ^i in Σ_3, both spaces being referred to special coordinates, then from (ii) and from the relation between the two metrics given in (2), it follows that

$$\underset{4}{\lambda^i} = c\lambda^i. \tag{15}$$

The vector λ^i gives no information about the remaining component $\underset{4}{\lambda^0}$ of the 4-vector, but it follows from (i) that $\underset{4}{\lambda^0}$ is a scalar object, *i.e.* is invariant in passing from one observer to any other. This component is thus arbitrary, and will in each case be chosen in some way appropriate to the object under consideration.

Similarly, covariant components are given by $\underset{4}{\lambda_i} = c^{-1}\lambda_i$, and $\underset{4}{\lambda_0}$ is an arbitrary scalar. Again, if $\underset{4}{T^{\alpha\beta}}$ is the 4-space extension of a tensor object T^{ij} in 3-space, then $\underset{4}{T^{ij}} = c^2 T^{ij}$, while $\underset{4}{T^{i0}}$, $\underset{4}{T^{0i}}$ are arbitrary vectors in 3-space and $\underset{4}{T^{00}}$ is an arbitrary scalar.

The motion of a particle is now given completely by its path in 4-space, along which any convenient variable can be used as a parameter. Objects relating to the particle and its motion are defined along the path, and the simplest way of measuring change is to take the absolute derivative in 4-space. If Δ denotes the absolute differential in Σ_4, then the change in a vector $\underset{4}{\lambda^a}$ is measured by

$$\underset{4}{\Delta\lambda^a} = \underset{4}{d\lambda^a} + \underset{4}{\Gamma'^{\alpha}_{\beta\gamma}}\underset{4}{\lambda^\beta}\, dx^\gamma,$$

where $\underset{4}{\Gamma'^{\alpha}_{\beta\gamma}}$ are the Christoffel symbols in Σ_4. It can be seen from (2) that in coordinates τ, ζ^i the only non-zero symbols are $\underset{4}{\Gamma'^{i}_{jk}}$ and that these are equal to the symbols Γ'^{i}_{jk} in Σ_3. Hence in the special coordinate system we have

$$\underset{4}{\Delta\lambda^0} = \underset{4}{d\lambda^0}, \quad \underset{4}{\Delta\lambda^i} = \underset{4}{\delta\lambda^i}. \tag{16}$$

These equations relate the operators in Σ_3 and Σ_4.

† The figure 4 below a tensor denotes the extension to 4-space of a tensor already occurring in 3-space. A prime, in general, denotes the tensor in $\underset{4}{\Sigma}$ which is equivalent to the tensor without a prime in S_4. Thus, $\underset{4}{p^a}$, $\underset{4}{p'^a}$ are the momentum vectors in S_4 and $\underset{4}{\Sigma}$ respectively, $\underset{4}{F^a}$, $\underset{4}{F'^a}$ the force vectors, &c.

In a general coordinate system x^α, the public time τ is a function of coordinates, and if N'^α is the unit normal to the surface $\tau = constant$ at any point, then

$$N'^\alpha = \gamma^{\alpha\beta} \frac{\partial \tau}{\partial x^\beta}. \tag{17}$$

This follows from the fact that the components of N'^α in the special coordinate system are (1, 0, 0, 0), so that, from (2), the above equation is satisfied in one coordinate system; it is therefore true in any other system. From (16) we see that, in the special coordinate system,

$$\Delta N'^\alpha = 0. \tag{18}$$

Since this is a tensor equation, it holds in all other coordinate systems.

4. We now consider the dynamical objects *momentum*, *force* and *work* defined in [1].

Momentum. From (7) and (15) we see that if $\underset{4}{p'^\alpha}$ is the momentum vector in Σ_4, then, in special coordinates, $\underset{4}{p'^i} = cp^i = Mc \frac{d\zeta^i}{dT'}$. Since momentum is proportional to velocity, which is tangent to the path in 3-space, it is evident that the most simple extension to 4-space is to take momentum as tangent to the path in this space also. This is already satisfied by three components and it is satisfied completely if we take $\underset{4}{p'^0} = Mc \frac{d\tau}{dT'} = mc\theta\phi$ in special coordinates. We now have, in *any* coordinate system,

$$\underset{4}{p'^\alpha} = Mc \frac{dx^\alpha}{dT'}. \tag{19}$$

From (14) we find that the momentum 4-scalar is

$$\left(\gamma_{\alpha\beta} \underset{4}{p'^\alpha} \underset{4}{p'^\beta}\right)^{\frac{1}{2}} = mc\eta^{\frac{1}{2}}\theta\phi. \tag{20}$$

This quantity thus has the constant value mc in the W-system of mechanics.

Force. If $\underset{4}{F'^\alpha}$ is the force vector in Σ_4, then from (8), (15) and (16) we see that

$$\underset{4}{F'^i} = cF'^i = c \frac{\delta p^i}{dT'} = \frac{\Delta}{dT'} \left(\underset{4}{p'^i}\right)$$

in special coordinates. It is now clear that the most suitable definition of force in Σ_4 is, in any coordinate system,

$$\underset{4}{F}'^{\alpha} = \frac{\Delta}{dT'}\left(\underset{4}{p}'^{\alpha}\right). \tag{21}$$

This means that, in special coordinates,

$$\underset{4}{F}'^{0} = mc\,\frac{d}{dT'}(\theta\phi). \tag{22}$$

It can easily be verified that the components of force are all zero when a particle is moving freely, a property which is obviously desirable.

Some useful identities are satisfied by force in 4-space. First, since $N'^{\alpha} = (1, 0, 0, 0)$ in special coordinates, we see from (22) that, in any coordinate system,

$$\gamma_{\alpha\beta}\underset{4}{F}'^{\alpha}N'^{\beta} = mc\,\frac{d}{dT'}(\theta\phi). \tag{23}$$

Secondly we have

$$\gamma_{\alpha\beta}\underset{4}{F}'^{\alpha}\frac{dx^{\beta}}{dT'} = \frac{1}{2Mc}\frac{d}{dT'}\left(\gamma_{\alpha\beta}\underset{4}{p}'^{\alpha}\underset{4}{p}'^{\beta}\right) = \frac{mc}{2\phi}\frac{d}{dT'}(\eta\theta^2\phi^2). \tag{24}$$

From (13) and (24) it at once follows that *in the W-system, force is always orthogonal to the path in* 4*-space.* This incidentally proves that, in this system, *force is a space-like vector.* For a tangent to the path of a particle is time-like, and hence force is orthogonal to a time-like vector in 4-space with signature -2.

By eliminating $d\eta/dT'$ between (23) and (24) we see that, in any system of mechanics, force is always orthogonal to some vector depending at most upon position and velocity. In particular we have

M-system: $$\gamma_{\alpha\beta}F'^{\alpha}\left(N'^{\beta} - 2\eta^{-\frac{1}{2}}\frac{dx^{\beta}}{d\sigma}\right) = 0$$

showing that in this case force is orthogonal to the vector

$$f^{\alpha} = N'^{\alpha} - 2\eta^{-\frac{1}{2}}\frac{dx^{\alpha}}{d\sigma}. \tag{25}$$

It is easily verified that $\gamma_{\alpha\beta}f^{\alpha}f^{\beta} = 1$, and hence f^{α} is a unit time-like vector. Here again, therefore, force is a space-like vector.

Work. From the definition of work in the W-system it follows that the rate of change of work is, in 4-space, the product of force and relative

velocity, *i.e.* velocity relative to the sub-stratum. From (19) we see that the velocity vector in 4-space is $cdx^\alpha/d\tau$ in the W-system. The velocity of the sub-stratum is therefore $(c, 0, 0, 0)$ in special coordinates, *i.e.* cN'^α generally, and the relative velocity is $c\dfrac{dx^\alpha}{d\tau} - cN'^\alpha$. Since this and the force vector are space-like, their product is taken with $-\gamma_{\alpha\beta}$, and we find from (23) and (24), as in [1],

$$W\text{-system}: \qquad \frac{dW}{d\tau} = -\gamma_{\alpha\beta} F'^\alpha \left(c\frac{dx^\beta}{d\tau} - cN'^\beta\right) = \frac{dE}{d\tau}. \tag{26}$$

The treatment given by Milne is essentially different, for it is not based upon the 3-space definition of work as the product of force and distance-displacement. In the M-system, work has only a 4-space definition, for its rate of change is the product of force and relative velocity, where the latter is now defined as relative momentum divided by relative mass. The momentum is now $Mc\,dx^\alpha/d\sigma$, and the momentum of a similar particle in the sub-stratum is $(mc, 0, 0, 0)$ in special coordinates, *i.e.* mcN'^α generally. The relative velocity defined by Milne is therefore

$$\frac{1}{M}\left(Mc\frac{dx^\alpha}{d\sigma} - mcN'^\alpha\right) = c\frac{dx^\alpha}{d\sigma} - c\eta^{\frac{1}{2}} N'^\alpha,$$

and we have, from (23) and (24),

$$M\text{-system}: \qquad \frac{dW}{d\sigma} = -\gamma_{\alpha\beta} F'^\alpha \left(c\frac{dx^\beta}{d\sigma} - c\eta^{\frac{1}{2}} N'^\beta\right) = \frac{dE}{d\sigma}. \tag{27}$$

This calculation, in other coordinates, was in fact used by Milne to determine the relation between m and M. A disadvantage of this treatment is that the increment of work is not exactly the product of the force and an actual displacement, and the similarity between this and the classical concept of work is artificial. The difference is emphasised when it is desired to define a conservative field of force along classical lines.

5. *Moments.* We now consider the moments of vectors about the origin, for it may be of interest to show how the idea of moments, essentially 3-dimensional in definition, can be extended to 4-space. This is made possible by the fact that moments are calculated by means of a moment tensor m'_{ij} in Σ_3, which can be extended to a tensor $\underset{4}{m'_{\alpha\beta}}$ in Σ_4, the moment of a vector $\underset{4}{\lambda^\alpha}$ being the covariant vector $\underset{4}{m'_{\alpha\beta}}\underset{4}{\lambda^\beta}$. To agree with moments in Σ_3 it is required that, if $\underset{4}{\lambda^\alpha}$ is a 4-vector corresponding

to λ^i in Σ_3, then the moment $\underset{4}{m'_{\alpha\beta}}\underset{4}{\lambda^\beta}$ in Σ_4 must correspond to the moment $m'_{ij}\lambda^j$ in Σ_3, *i.e.*, in special coordinates,

$$\underset{4}{m'_{i\beta}}\underset{4}{\lambda^\beta} = \frac{1}{c}\, m'_{ij}\lambda^j, \tag{28}$$

where $\underset{4}{\lambda^i} = c\lambda^i$ and $\underset{4}{\lambda^0}$ is arbitrary. We thus have

$$\underset{4}{m'_{ij}} = c^{-2}\, m'_{ij}, \quad \underset{4}{m'_{i0}} = 0,$$

and this determines $\underset{4}{m'_{\alpha\beta}}$ if it is taken to be skew-symmetric as in 3-space. To find an expression in terms of any coordinates, we note from (10) that, in special coordinates,

$$\underset{4}{m'_{ij}} = -c^{-2}\sqrt{(|k|)}\, e_{ijk}\, P'^k.$$

We can write $e_{ijk} = e_{0ijk}$, where $e_{\alpha\beta\gamma\delta}$ is zero if two suffixes are equal and otherwise is ± 1 according as $\alpha\beta\gamma\delta$ is an even or odd permutation of 0123. The vector P'^k can also be replaced by a 4-vector defined in special coordinates by†

$$\underset{4}{P'^i} = cP'^i, \quad \underset{4}{P'^0} = 0.$$

This vector can be transformed to any coordinate system and will be written $\underset{4}{P'^\alpha}$ generally. Now, since $|k| = -c^6|\gamma|$ from (2),

$$\underset{4}{m'_{ij}} = -\sqrt{(-|\gamma|)}\, e_{0ij\delta}\, \underset{4}{P'^\delta} = -\epsilon'_{0ij\delta}\, \underset{4}{P'^\delta},$$

where $\epsilon'_{\alpha\beta\gamma\delta}$ is the tensor $\sqrt{(-|\gamma|)}\, e_{\alpha\beta\gamma\delta}$ in Σ_4. Hence, since $\underset{4}{m'_{\alpha 0}} = \underset{4}{m'_{0\alpha}} = 0$ and $N'^\alpha = (1, 0, 0, 0)$ in special coordinates, we have finally

$$\underset{4}{m'_{\alpha\beta}} = -\epsilon'_{\alpha\beta\gamma\delta}\, N'^\gamma\, \underset{4}{P'^\delta}. \tag{29}$$

This is a tensor equation applying to any system of coordinates provided that $\underset{4}{P'^\alpha}$ is calculated correctly as the transform of the vector $(0, c\zeta^i)$ in special coordinates.

To find the 4-space equivalent of the angular equations of motion, we see from (28) and the covariant form of (17) that, in special coordinates,

$$\underset{4}{m'_{i\beta}}\,\underset{4}{F'^\beta} = \frac{1}{c}\, m'_{ij}\, F'^j = \frac{1}{c}\,\frac{\delta}{dT'}\,(m'_{ij}\, p^j) = \frac{\Delta}{dT'}\left(\underset{4}{m'_{i\beta}}\,\underset{4}{p'^\beta}\right).$$

† The value of $\underset{4}{P'^0}$ does not affect the tensor $\underset{4}{m'_{\alpha\beta}}$.

Also, in these coordinates, $\underset{4}{m'_{0\beta}} = 0$, so that, from (17),

$$\underset{4}{m'_{0\beta}}\,\underset{4}{F'^{\beta}} = 0 = \frac{\Delta}{dT'}\left(\underset{4}{m'_{0\beta}}\,\underset{4}{p'^{\beta}}\right).$$

Thus finally, in any system of coordinates,

$$\underset{4}{m'_{\alpha\beta}}\,\underset{4}{F'^{\beta}} = \frac{\Delta}{dT'}\left(\underset{4}{m'_{\alpha\beta}}\,\underset{4}{p'^{\beta}}\right), \tag{30}$$

i.e. the moment of force is still the absolute rate of change of the angular momentum.

6. We now transform from space Σ_4 with time τ to space S_4 with time t. Our method is to adopt a general coordinate system x^α in S_4, the metric being

$$ds^2 = g_{\alpha\beta}\,dx^\alpha\,dx^\beta, \tag{31}$$

and to consider the objects in Σ_4 described in terms of the *same coordinates*. The transformation from one space to the other is then a simple matter, since we have already described objects in terms of general coordinates in Σ_4. With this choice of coordinates, since $ds = (t/t_0)\,d\sigma$, we have

$$g_{\alpha\beta} = \left(\frac{t}{t_0}\right)^2 \gamma_{\alpha\beta}, \quad g^{\alpha\beta} = \left(\frac{t_0}{t}\right)^2 \gamma^{\alpha\beta}. \tag{32}$$

If therefore an object is described by vectors λ^α in S_4 and λ'^α in Σ_4, then

$$\lambda^\alpha = \frac{t_0}{t}\,\lambda'^\alpha. \tag{33}$$

Other tensors are similarly related.

In a general coordinate system, public time t is a function of coordinates, and, as in Σ_4, the unit normal to the surface $t = constant$ in S_4 is N^α, where

$$N^\alpha = g^{\alpha\beta}\,\frac{\partial t}{\partial x^\beta}. \tag{34}$$

From (3) and (32) we see by direct transformation that the vectors N^α and N'^α are related by

$$N^\alpha = \frac{t_0}{t}\,N'^\alpha. \tag{35}$$

The absolute differential in S_4 will be denoted by D, so that

$$D\lambda^\alpha = d\lambda^\alpha + \underset{4}{\Gamma}{}^\alpha_{\beta\gamma}\lambda^\beta dx^\gamma, \tag{36}$$

etc., where $\underset{4}{\Gamma}{}^\alpha_{\beta\gamma}$ are the Christoffel symbols in S_4. To relate D and Δ, we find from (32) that

$$\underset{4}{\Gamma}{}^\alpha_{\beta\gamma} = \underset{4}{\Gamma}{}'^\alpha_{\beta\gamma} + \frac{1}{t}\left(\delta_\beta{}^\alpha \frac{\partial t}{\partial x^\gamma} + \delta_\gamma{}^\alpha \frac{\partial t}{\partial x^\beta} - g_{\beta\gamma}g^{\alpha\delta}\frac{\partial t}{\partial x^\delta}\right). \tag{37}$$

From (34) we can therefore write

$$\frac{1}{t}\Delta(t\lambda^\alpha) = D\lambda^\alpha - \frac{1}{t}(g_{\beta\gamma}\lambda^\beta N^\gamma)\,dx^\alpha + \frac{1}{t}(g_{\beta\gamma}\lambda^\beta dx^\gamma)\,N^\alpha. \tag{38}$$

From (18), (34) and (35) it follows that

$$D(tN^\alpha) = dx^\alpha. \tag{39}$$

7. In changing from time τ to time t it must be remembered from results in 3-space that some objects, such as momentum, are unaltered, while for others it is convenient to take slightly different definitions. Such differences are shown in (6), (8), (9) and (10), for it may be recalled that a given object has the same components in both S_3 and Σ_3. Thus $F'^i = (t/t_0)F^i$ shows that force in τ-time is defined to be t/t_0 times force in t-time, etc.

Momentum. From (19) and (33), the momentum vector in S_4 is $(t_0/t)\,Mc\,(dx^\alpha/dT')$, *i.e.* from (6),

$$\underset{4}{p}{}^\alpha = Mc\frac{dx^\alpha}{dT}. \tag{40}$$

Force. Since $F^i = (t_0/t)F'^i$ in 3-space, it follows from (33) that the force vector in S_4 is

$$\underset{4}{F}{}^\alpha = \left(\frac{t_0}{t}\right)^2 \underset{4}{F}{}'^\alpha = \left(\frac{t_0}{t}\right)^2 \frac{\Delta}{dT'}\left(\underset{4}{p}{}'^\alpha\right) = \frac{t_0}{t}\frac{\Delta}{dT}\left(\frac{t}{t_0}\underset{4}{p}\right),$$

and from (38) this can be written

$$\underset{4}{F}{}^\alpha = \frac{D}{dT}\left(\underset{4}{p}{}^\alpha\right) - \frac{1}{t}\left(g_{\beta\gamma}\underset{4}{p}{}^\beta N^\gamma\right)\frac{dx^\alpha}{dT} + \frac{1}{t}\left(g_{\beta\gamma}\underset{4}{p}{}^\beta\frac{dx^\gamma}{dT}\right)N^\alpha.$$

From (34) and (40) we see that

$$g_{\beta\gamma}\underset{4}{p}{}^\beta N^\gamma = Mc\frac{dt}{dT} = Mc\theta,$$

and from (31),

$$g_{\beta\gamma} \underset{4}{p}{}^{\beta} \frac{dx^{\gamma}}{dT} = Mc\left(\frac{ds}{dT}\right)^2 = Mc\eta\theta^2.$$

Hence, finally,

$$\underset{4}{F}{}^{\alpha} = \frac{D}{dT}\left(\underset{4}{p}{}^{\alpha}\right) + \frac{Mc\theta}{t}\left(\eta\theta N^{\alpha} - \frac{dx^{\alpha}}{dT}\right). \tag{41}$$

The identities (23) and (24) become

$$g_{\alpha\beta} \underset{4}{F}{}^{\alpha} N^{\beta} = mc \frac{d}{dT}(\theta\phi), \tag{42}$$

$$g_{\alpha\beta} \underset{4}{F}{}^{\alpha} \frac{dx^{\beta}}{dT} = \frac{mc}{2\phi} \frac{d}{dT}(\eta\theta^2\phi^2). \tag{43}$$

In the W-system, force is orthogonal to the path in S_4, as in Σ_4.

Moments. Since $P^i = (t/t_0)\, P'^i$ in (9), the most suitable position vector in S_4 is, from (33),

$$\underset{4}{P}{}^{\alpha} = \underset{4}{P'}{}^{\alpha}. \tag{44}$$

The components of $\underset{4}{P'}{}^{\alpha}$ are 0, $c\zeta^i$ in coordinates τ, ζ^i, and from (3) it follows that, in coordinates t, z^i, the vector $\underset{4}{P}{}^{\alpha}$ is given by

$$\underset{4}{P}{}^{0} = 0, \quad \underset{4}{P}{}^{i} = cz^i. \tag{45}$$

From (10) and from the appropriate equation similar to (33) it follows that the moment tensor in S_4 is

$$\underset{4}{m}_{\alpha\beta} = \left(\frac{t}{t_0}\right)^3 \underset{4}{m'}_{\alpha\beta} = -\left(\frac{t}{t_0}\right)^3 \sqrt{(-|\gamma|)}\, e_{\alpha\beta\gamma\delta} N'^{\gamma} \underset{4}{P'}{}^{\delta}.$$

From (32), $|\gamma| = (t_0/t)^8 |g|$, and writing $\epsilon_{\alpha\beta\gamma\delta} = \sqrt{(-|g|)}\, e_{\alpha\beta\gamma\delta}$, we have finally, from (35) and (44),

$$\underset{4}{m}_{\alpha\beta} = -\epsilon_{\alpha\beta\gamma\delta} N^{\gamma} \underset{4}{P}{}^{\delta}. \tag{46}$$

This gives $\underset{4}{m}_{\alpha\beta}$ in tensor form in S_4, as required.

From the covariant equation similar to (38), it can now easily be verified that (30) is transformed into

$$\frac{1}{t} \underset{4}{m}_{\alpha\beta} \underset{4}{F}{}^{\beta} = \frac{D}{dT}\left(\frac{1}{t} \underset{4}{m}_{\alpha\beta} \underset{4}{p}{}^{\beta}\right). \tag{47}$$

This, then, is the relation between angular momentum and moment of force in S_4 with t-time.

8. As an example of the way in which the general formulae found above can be applied to other coordinate systems, we shall express them in terms of the coordinates used by Milne. It will then be seen that the M-system is in fact the scheme of mechanics developed by Milne. The constant K now takes the value -1, since this is the only case considered by Milne, and for the same reason we restrict ourselves to the M-system, although any other system could be transformed equally well.

The transformation from special coordinates t, z^i to Milne's coordinates t', x, y, z is {[1], equation (21)}

$$t' = t\,\frac{1+\frac{1}{4}|z|^2}{1-\frac{1}{4}|z|^2}, \quad \mathbf{P} = (x, y, z) = \frac{ctz}{1-\frac{1}{4}|z|^2}, \tag{48}$$

and the metric of S_4 becomes

$$ds^2 = dt'^2 - \frac{1}{c^2}\,(dx^2+dy^2+dz^2). \tag{49}$$

If we write $\mathbf{V} = d\mathbf{P}/dt'$, the invariants used by Milne are

$$\left.\begin{aligned} X &= t'^2 - \frac{1}{c^2}\,\mathbf{P}^2 = t^2, \qquad Y = 1 - \frac{1}{c^2}\,\mathbf{V}^2 = \left(\frac{ds}{dt'}\right)^2, \\ Z &= t' - \frac{1}{c^2}\,\mathbf{P}\,.\,\mathbf{V} = t\,\frac{dt}{dt'}, \quad \xi = \frac{Z^2}{XY} = \frac{1}{\eta}. \end{aligned}\right\} \tag{50}$$

In the new coordinates the vector N^α is given by (34), and we get

$$N^\alpha = \frac{1}{t}\,(t', x, y, z). \tag{51}$$

It is useful to introduce the unit normal n^α to the surface $t' = constant$. This has components (1, 0, 0, 0) in Milne's coordinates and is therefore $g^{\alpha\beta}(\partial t'/\partial x^\beta)$ generally. Applying this formula to the special coordinate system t, z^i, we deduce from (48) that

$$n^0 = \frac{t'}{t}, \quad n^i = -\frac{z^i}{t}.$$

It now follows from (45), by considering their components in special coordinates, that the three vectors P^α, N^α and n^α are connected by the invariant equation

$$P^\alpha = ct'\,N^\alpha - ctn^\alpha. \tag{52}$$

9. The Euclidean space with coordinates x, y, z and metric

$$dx^2+dy^2+dz^2$$

is regarded as the "private" space belonging to the observer at the origin, and a 4-vector $\underset{4}{\lambda^\alpha}$ in Milne's coordinates is now described as having a time component $\underset{4}{\lambda^0}$ and a spatial part given by the Euclidean vector

$$\boldsymbol{\lambda} = \frac{1}{c}\left(\underset{4}{\lambda^1}, \underset{4}{\lambda^2}, \underset{4}{\lambda^3}\right). \tag{53}$$

Similarly, a vector $\underset{4}{\lambda_\alpha}$ in S_4 has time component $\underset{4}{\lambda_0}$ and spatial part $\boldsymbol{\lambda} = c\left(\underset{4}{\lambda_1}, \underset{4}{\lambda_2}, \underset{4}{\lambda_3}\right)$. These conventions follow from the form of the metric (49).

In S_4 the moment about O of a vector $\underset{4}{\lambda^\alpha}$ is the covariant vector which, from (52), can be written $\underset{4}{m_{\alpha\beta}}\underset{4}{\lambda^\beta} = ct\sqrt{(-|g|)}\,e_{\alpha\beta\gamma\delta}N^\gamma n^\delta \underset{4}{\lambda^\beta}$. In Milne's coordinates, $|g| = -c^{-6}$, $N^\alpha = t^{-1}$ (t', x, y, z) and $n^\alpha = (1, 0, 0, 0)$. It at once follows that the time component of the moment is zero, and the spatial part is the Euclidean vector $\mathbf{P}\times\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is the spatial part of $\underset{4}{\lambda^\alpha}$. Our general definition of moment thus gives the classical form in Milne's coordinates.

Momentum. Remembering that $dT = ds$ in the M-system, we see from (40) that momentum has time component p^0 and spatial part $\mathbf{p}$, where

$$p^0 = Mc\frac{dt'}{ds} = \frac{Mc}{Y^{\frac{1}{2}}}, \quad \mathbf{p} = M\frac{d\mathbf{P}}{ds} = M\frac{\mathbf{V}}{Y^{\frac{1}{2}}}. \tag{54}$$

Force. Writing $\theta = \eta^{-\frac{1}{2}}$ in (41) and noting from (49) that the absolute derivative is now the ordinary derivative, we see that force has time component F^0 and spatial part $\mathbf{F}$, where

$$F^0 = \frac{1}{Y^{\frac{1}{2}}}\frac{dp^0}{dt'} + \frac{Mc}{X}\left(t' - \frac{Z}{Y}\right), \quad \mathbf{F} = \frac{1}{Y^{\frac{1}{2}}}\frac{d\mathbf{p}}{dt'} + \frac{M}{X}\left(\mathbf{P} - \frac{Z}{Y}\mathbf{V}\right). \tag{55}$$

The identities (42) and (43) can now be written

$$ct'\,F^0 - \mathbf{P}.\mathbf{F} = 2c^2\frac{Z}{Y}\frac{dM}{dt'}, \tag{56}$$

$$cF^0 - \mathbf{V}.\mathbf{F} = c^2\frac{dM}{dt'}. \tag{57}$$

Formulae equivalent to these are given by Milne in [2].

Moments. From the remarks concerning moments given above, it follows from (47) that angular momentum and force now satisfy

$$\frac{1}{X^{\frac{1}{2}}}\mathbf{P}\times\mathbf{F}=\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt'}\left(\frac{1}{X^{\frac{1}{2}}}\mathbf{P}\times\mathbf{p}\right). \tag{58}$$

10. In the foregoing sections we have been concerned only with t-time since Milne's coordinates then require the formulae in 4-space. With time τ, Milne's results are given almost immediately by our formulae in 3-space Σ_3, since these formulae apply to any coordinate system in the space, and in particular to the coordinates used by Milne. These are λ, θ, ϕ where θ and ϕ are polar angles and λ is related to the special coordinates ζ^i by

$$\lambda=2ct_0\tanh^{-1}\frac{|\zeta|}{2ct_0}, \tag{59}$$

since $K=-1$. The metric of Σ_3 now becomes

$$d\epsilon^2=d\lambda^2+\left(ct_0\sinh\frac{\lambda}{ct_0}\right)^2(d\theta^2+\sin^2\theta d\phi^2), \tag{59'}$$

and Milne's results (see [2] and [3]) are given by our formulae with the appropriate fundamental tensor.

11. We now turn to the problem of extending to relativity mechanics the definitions and properties of conservative fields of force and potential. We first give a brief account of Milne's procedure, and then suggest an alternative definition.

Generalising to 4-space the classical result that force in a conservative field is the gradient of some potential function, Milne first considered the 4-vector equivalent to† $(m/c)g^{\alpha\beta}(\partial\chi/\partial x^\beta)$. This, however, contradicts the force identities, so a term proportional to velocity is added, giving the form

$$\underset{4}{F}{}^\alpha=\frac{m}{c}g^{\alpha\beta}\frac{\partial\chi}{\partial x^\beta}+a\frac{dx^\alpha}{ds}.$$

The identity (42) now gives a owing to the assumption made by Milne {[2], (16′)} that potential should satisfy $(\partial\chi/\partial x^\alpha)N^\alpha=0$. In the M-system, this gives $a=2cdM/ds$, and the equation relating potential and force in

† The force-potential for a particle of proper mass m is, in our notation, $m\chi$.

a conservative field is now

$$F^{\alpha}_{4} = \frac{m}{c} g^{\alpha\beta} \frac{\partial \chi}{\partial x^{\beta}} + 2c \frac{dM}{ds} \frac{dx^{\alpha}}{ds} = \frac{m}{c} g^{\alpha\beta} \frac{\partial \chi}{\partial x^{\beta}} - \frac{2m}{c} \frac{d\chi}{ds} \frac{dx^{\alpha}}{ds}. \tag{60}$$

In Milne's coordinates this can be written

$$F^{0} = \frac{m}{c} \frac{\partial \chi}{\partial t'} + \frac{2c}{Y} \frac{dM}{dt'}, \quad \mathbf{F} = -m \frac{\partial \chi}{\partial \mathbf{P}} + \frac{2}{Y} \frac{dM}{\partial t'} \mathbf{V}. \tag{61}$$

It can be verified that these equations yield $E + m\chi = constant$ along a path in the field of force. Also the work done by the field when a particle is constrained to move along a prescribed path from one point to another in 4-space is independent of the path. Here we take work to be force $\times$ displacement, where displacement is relative velocity $\times ds$ or $d\sigma$ (see § 4), and use the second form of (60) giving field-force in terms of position and velocity. This treatment is, however, essentially 4-dimensional, for it can be verified that Milne's definition of potential yields no corresponding result in 3-space. This, I believe, is a disadvantage, but there is another far more significant, for it is easily seen that, with the above definition, motion in a conservative field does not satisfy a variational equation.

12. An alternative definition of a conservative field of force arises in the W-system, but it should be remarked that there is no *a priori* reason for choosing one particular definition. In classical gravitational and electromagnetic theories potential was introduced as a convenience to simplify descriptions of certain systems of motions, but in extending the idea to relativity we can define a conservative field of force as we please. A natural preference may arise when, as is hoped will be the case, relativity methods are adapted to deal with local problems, but at present the only distinction between different definitions is in the way the results compare with classical results.

In the W-system we can define a conservative field of force so that the work done by the field on a particle is independent of the path in 3-space and independent also of the time taken. This definition implies the existence of a potential function χ, which is a function of the ζ's (or z's) but is independent of τ (or t), the work done on a particle m in any small displacement being $dW = -md\chi$. Since $dW = k_{ij} F'^{i} d\zeta^{j}$ in Σ_3, it follows that force and potential must satisfy the equations

$$\frac{\partial \chi}{\partial \tau} = 0, \quad F'^{i} = -mk^{ij} \frac{\partial \chi}{\partial \zeta^{j}} \tag{62}$$

in τ-time, and the equations

$$\frac{\partial \chi}{\partial t} = 0, \quad F^i = -\frac{m}{ct} k^{ij} \frac{\partial \chi}{\partial z^j} \tag{63}$$

in t-time. In 4-space, the corresponding equations are

$$\frac{\partial \chi}{\partial x^a} N'^a = 0, \quad \frac{\partial \chi}{\partial x^a} N^a = 0, \tag{64}$$

τ-time:
$$\underset{4}{F}'^a = \frac{m}{c} \gamma^{a\beta} \frac{\partial \chi}{\partial x^\beta} + c \frac{dM}{d\tau} N'^a; \tag{65}$$

t-time:
$$\underset{4}{F}^a = \frac{m}{c} g^{a\beta} \frac{\partial \chi}{\partial x^\beta} + c \frac{dM}{dt} N^a. \tag{66}$$

The similarity between (62) and the classical equations is apparent. Further, since $dW = dE$, where $E = Mc^2 = mc^2 \eta^{-\frac{1}{2}}$, the equations of motion have a first integral

$$E + m\chi = \text{constant}. \tag{67}$$

Again, we see from [1], equation (93), that the motion is now given by a variational equation

$$\delta \int (E^* - m\chi)\, d\tau = 0, \tag{68}$$

where $E^* = -mc^2 \eta^{\frac{1}{2}}$. It appears, therefore, that this definition of a conservative field gives all the desired results.

We see that classical kinetic energy is now replaced by E in some formulae and E^* in others; each approximates to const. $+\frac{1}{2}mv^2$ when v/c is small.

It may be of interest to remark that the above equations in a conservative field can be obtained by requiring that the equations of motion have (67) as a first integral and also satisfy a variational equation $\delta \int L\, d\tau = 0$, where L is a function of position and scalar velocity, *i.e.* $L = L(\tau, \zeta, \eta)$. It can then be shown that L must be of the form $E^* - m\chi$, as in (68), and equations (62) follow at once.

It should be noted that a conservative field as defined above differs essentially from the field defined by Milne; it is not now simply a question of different descriptions of the same thing. The equations of motion in the conservative field discussed here can, of course, be described in the M-system, but equations (62) then lose their simplicity. In terms of

Milne's coordinates, the equations in the M-system become

$$\left.\begin{aligned} F^0 &= \frac{m\xi}{c}\frac{\partial\chi}{\partial t'}+\frac{c\xi}{Z}\frac{dM}{dt'}\left(t'+\frac{X}{Z}\right), \\ \mathbf{F} &= -m\xi\frac{\partial\chi}{\partial\mathbf{P}}+\frac{\xi}{Z}\frac{dM}{dt'}\left(\mathbf{P}+\frac{X}{Z}\mathbf{V}\right). \end{aligned}\right\} \tag{69}$$

These equations replace (61) if our definition of a conservative field is adopted.

13. A simple example of a conservative field of force is given by a potential which depends only upon the distance from the origin. Working in τ-time and with special coordinates ζ^i, and writing $r=|\zeta|$, we have $\chi=\chi(r)$. If now we adopt the W-system and the definition of potential described above, it is seen from (62) that the force is central†, *i.e.* $F'^i\propto\zeta^i$. This leads to an angular momentum integral, for from [1], equation (102), we find $\nu m'_{ij}p^j=mh_i$, where

$$\nu=1+\frac{Kr^2}{4c^2t_0{}^2}, \quad r^2=(\zeta^1)^2+(\zeta^2)^2+(\zeta^3)^2, \tag{70}$$

and the h's are constants. Since now $|k_{ij}|=\nu^{-6}$ and $p^i=m\eta^{-\frac{1}{2}}(d\zeta^i/d\tau)$, this integral becomes

$$e_{ijk}\zeta^j\frac{d\zeta^k}{d\tau}=\nu^2\eta^{\frac{1}{2}}h_i.$$

This at once indicates that each path lies in a surface of geodesics through the origin, corresponding to a plane in Euclidean space. For convenience this surface can be taken to be $\zeta^3=0$, and we can refer to polar coordinates‡ r, ϕ, where

$$\zeta^1=r\cos\phi, \quad \zeta^2=r\sin\phi, \quad d\epsilon^2=\nu^{-2}(dr^2+r^2\dot{d}\phi^2).$$

The angular momentum integral now becomes

$$r^2\frac{d\phi}{d\tau}=h\nu^2\eta^{\frac{1}{2}}, \tag{71}$$

† This result does not follow from Milne's definition of potential.

‡ These coordinates agree very closely with Euclidean polars for distances from the origin such that r is small compared with ct_0.

where h is a constant, and writing $u = 1/r$, $u' = du/d\phi$, etc., we can easily verify from the form of η that the integral is equivalent to

$$\frac{1}{\eta} = 1 + \frac{h^2 \nu^2}{c^2} (u'^2 + u^2). \tag{72}$$

The path equation is now given by the energy integral

$$m\chi + mc^2 \eta^{-\frac{1}{2}} = \text{constant},$$

and we have

$$\chi + c^2 \left\{ 1 + \frac{h^2 \nu^2}{c^2} (u'^2 + u^2) \right\}^{\frac{1}{2}} = \text{constant}. \tag{73}$$

The normal equation, which is frequently more convenient, is found by differentiating (73) with respect to ϕ and dividing by u'. This becomes

$$u'' + u - \frac{K(u'^2 + u^2)}{2c^2 t_0{}^2 \nu u^3} + \frac{1}{h^2 \nu^2} \frac{d\chi}{du} \left\{ 1 + \frac{h^2 \nu^2}{c^2} (u'^2 + u^2) \right\}^{\frac{1}{2}} = 0, \tag{74}$$

which, we observe, reduces to the classical orbit equation when K and $1/c^2$ are neglected.

14. *Conclusions.* We have now explored thoroughly the elements of various schemes of mechanics, both in 3-space and 4-space, and with both t-time and τ-time, and we are in a position to make comparisons and draw conclusions. What can usefully be decided is, which is the best scheme of mechanics, which map (3- or 4-space) gives the best picture, and when we should refer to t-time and when to τ-time. In all these comparisons we mean by "best" the most simple mathematically and that which shows the closest analogy to classical mechanics.

In the first place it is clearly seen that there is, in general, no advantage in using a 4-space map. With τ-time the formulae in 4-space are no simpler or more significant than in 3-space, and with t-time, whatever coordinates are used, the formulae in 4-space are more complicated. It is advisable to keep the 4-space form in mind in case we wish to use coordinates other than those of public time and space, but it is doubtful whether any dynamical problem will make this necessary. The above general conclusion is at first surprising, for it means replacing the more recent pictures of mechanics by the earlier ones, using 3-dimensional space with time as a parameter. We observe, however, that the recent advances are not ignored and that their results are incorporated in our mechanics in such a way that we can retain the classical mental pictures and use them as guides to what we expect to discover. From the many formulae

constructed in the present and preceding papers it is evident that a simple principle of covariance will give immediately the relativistic analogues of most of the classical formulae.

Secondly, we have seen that even in public 3-space there are many possible schemes of mechanics, characterized by functions $\theta(\eta)$ and $\phi(\eta)$. It appears, however, that the useful classical idea of work can be taken over to relativity only if $\theta \equiv 1$ and $\phi \equiv \eta^{-\frac{1}{2}}$, *i.e.* in the W-system. In support of this system it is seen that the time interval for measuring rate of change is simply the interval $d\tau$ (or dt) of public time, and that relative mass is $M = m\eta^{-\frac{1}{2}}$, a form well established in relativity. Further, it is possible to define potential and a conservative field of force as in classical mechanics so that they have all the desired properties.

Lastly, we note that there is no question of choice when referring to t- or τ-time, since each has its own importance. Kinematic, or t-time, is significant in that it is the only scale of time which requires no dimensional constants in its derivation, and it is therefore to be used when any fundamental question of dimensions arises. On the other hand, dynamic, or τ-time, is significant in that it agrees most closely with the time appearing in the equations of classical mechanics, *i.e.* our relativistic formulae are most similar to the classical formulae when τ-time is used. This scale of time should therefore be used when we discuss the relativistic analogues of formal classical results, such as those connected with gravitational theory.

References.

1. Walker, A. G., "Relativistic mechanics", I & II, *Proc. London Math. Soc.* (2), 46 (1940), 113.
2. Milne, E. A. "Kinematics, dynamics, and the scale of time", II, *Proc. Royal Soc.* (A), 159 (1937), 171.
3. ————, "Kinematics, dynamics, and the scale of time", III, *Proc. Royal Soc.* (A), 159 (1937), 526.

Department of Pure Mathematics,
University of Liverpool.

A METHOD FOR THE CALCULATION OF THE ZETA-FUNCTION

By A. M. TURING.

[Received 7 March, 1939.—Read 16 March, 1939.]

An asymptotic series for the zeta-function was found by Riemann and has been published by Siegel*, and applied by Titchmarsh† to the calculation of the approximate positions of some of the zeros of the function. It is difficult to obtain satisfactory estimates for the remainders with this asymptotic series, as may be seen from the first of these two papers of Titchmarsh, unless t is very large. In the present paper a method of calculation will be described, which, like the asymptotic formula, is based on the approximate functional equation; it is applicable for all values of s. It is likely to be most valuable for a range of t where t is neither so small that the Euler-Maclaurin summation method can be used (*e.g.* $t > 30$) nor large enough for the Riemann-Siegel asymptotic formula (*e.g.* $t < 1000$).

Roughly speaking, the method is to use the approximate functional equation for the zeta-function, with the remainder expressed as an integral, which for the moment we write as $\int_{-\infty}^{\infty} h(x)\,dx$. We approximate to the integral by the obvious sum $\sum_{k=-K}^{K} \frac{1}{\kappa} h\left(\frac{k}{\kappa}\right)$ and we find that, if certain modifications are made in the "main series", this gives a remarkably accurate result; when the number of terms taken is $T = 2K+1$ the error is of the order of magnitude of $e^{-\frac{1}{2}\pi T}$. The theta-functions give another case of this phenomenon. We have the identity

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 x} = \frac{1}{\sqrt{x}} \sum_{n=-\infty}^{\infty} e^{-\pi n^2/x}$$

* C. L. Siegel, "Über Riemanns Nachlass zur analytischen Zahlentheorie", *Quell. Gesch. Math.*, B, 2 (1931), 45–80.

† E. C. Titchmarsh, "The zeros of the Riemann zeta-function", *Proc. Royal Soc.* (A), 151 (1935), 234–255; also 157 (1936), 261–263.

for any positive x. Hence

$$\left|\frac{1}{\kappa}\sum_{n=-\infty}^{\infty} e^{-\pi n^2/\kappa^2} - \int_{-\infty}^{\infty} e^{-\pi v^2}\,dv\right| = 2\sum_{n=1}^{\infty} e^{-\pi n^2\kappa^2} < \frac{2}{1-e^{-2\pi}}\,e^{-\pi\kappa^2} \quad (\text{if } \kappa \geqslant 1),$$

$$\left|\frac{1}{\kappa}\sum_{n=-K}^{K} e^{-\pi n^2/\kappa^2} - \int_{-\infty}^{\infty} e^{-\pi v^2}\,dv\right| < \frac{2}{1-e^{-2\pi}}\,e^{-\pi\kappa^2} + \frac{\kappa\, e^{-\pi K^2/\kappa^2}}{\pi K}$$

$$< e^{-\pi K}\left(\frac{2}{1-e^{-2\pi}} + \frac{1}{\pi}\right) \quad (\text{if } K = \kappa^2).$$

As we have proved it above, this inequality depends entirely on the special form of the function $e^{-\pi u^2}$. But we can also prove it in this way. We integrate the function $e^{-\pi u^2}/(1-e^{-2\pi i u\kappa})$ round the rectangle with the vertices $\pm R \pm i\kappa$. In the limit $R \to \infty$ we obtain, by the theorem of residues,

$$\left(\int_{-\infty-i\kappa}^{\infty-i\kappa} + \int_{\infty+i\kappa}^{-\infty+i\kappa}\right)\frac{e^{-\pi u^2}}{1-e^{-2\pi i u\kappa}}\,du = \sum_{k=-\infty}^{\infty}\frac{1}{\kappa}\,e^{-\pi k^2/\kappa^2},$$

i.e.

$$\sum_{k=-\infty}^{\infty}\frac{1}{\kappa}\,e^{-\pi k^2/\kappa^2} - \int_{-\infty-i\kappa}^{\infty-i\kappa} e^{-\pi u^2}\,du = \int_{-\infty-i\kappa}^{\infty-i\kappa}\frac{e^{-\pi u^2}\,e^{-2\pi i u\kappa}}{1-e^{-2\pi i u\kappa}}\,du + \int_{\infty+i\kappa}^{-\infty+i\kappa}\frac{e^{-\pi u^2}\,e^{2\pi i u\kappa}}{1-e^{2\pi i u\kappa}}\,du.$$

The path of integration on the left-hand side of the equation can be replaced by the real axis, while the right-hand side is less in modulus than

$$\frac{2}{1-e^{-2\pi\kappa^2}}\int_{-\infty}^{\infty}\left|\exp\left[-\pi(u-i\kappa)^2 - 2\pi i(u-i\kappa)\kappa\right]\right|du$$

$$\leqslant \frac{2}{1-e^{-2\pi\kappa^2}}\int_{-\infty}^{\infty}\exp\left(-\pi u^2 - \pi\kappa^2\right)du \leqslant \frac{2e^{-\pi\kappa^2}}{1-e^{-2\pi}} \quad (\text{if } \kappa \geqslant 1).$$

This argument can be used in more general cases, but in moving the path of integration we may encounter singularities of the integrand, which will modify the result.

We base our calculation on an integral representation of the zeta-function due to Riemann*.

1. *Evaluation of a definite integral.* Let

$$G(u) = \int_{0\nearrow 1}\frac{\exp\left(i\pi z^2 + 2\pi i u z\right)dz}{e^{i\pi z} - e^{-i\pi z}},$$

where $0 \nearrow 1$ signifies that the integration is along a line from $-\epsilon\infty$ to $\epsilon\infty$ cutting the real axis between 0 and 1, and $\epsilon = e^{\frac{1}{4}\pi i}$. We denote

* Siegel, *loc. cit.*, 24.

a similar integral over $-1 \nearrow 0$ by $G_1(u)$. Then, by a change of variables, we obtain

$$G_1(u) = G(u-1)\, e^{-2\pi i u},$$

and, by the theorem of residues,

$$G(u) = G_1(u) + 1.$$

Multiplying the identity

$$\frac{1}{e^{i\pi z} - e^{-i\pi z}} = e^{-i\pi z} + \frac{e^{-2\pi i z}}{e^{i\pi z} - e^{-i\pi z}}$$

by $\exp(i\pi z^2 + 2\pi i u z)$, and integrating over $0 \nearrow 1$, we obtain

$$G(u) = \epsilon\, e^{-i\pi(u-\frac{1}{2})^2} + G(u-1).$$

Combining these results we have

$$G(u) = \frac{1}{1-e^{-2\pi i u}} - \frac{e^{-i\pi u^2}}{e^{i\pi u} - e^{-i\pi u}}. \tag{1.1}$$

2. *An integral representation of the zeta-function.* We integrate $z^{-s}/(1-e^{-2\pi i z})$ round a curve L which may be taken to consist of a straight line from $i\infty$ to $\frac{1}{2}$, a semicircle from $\frac{1}{2}$ to $-\frac{1}{2}$ lying in the lower half-plane, and a straight line from $-\frac{1}{2}$ to $i\infty$. We define z^{-s} so that it has its usual value on the positive real axis, and is continuous except on the positive imaginary axis. The integral round any part of a circle with 0 as centre tends to 0 as the radius tends to infinity through appropriate values, provided that $\Re s > 1$, and therefore in this case we have

$$\int_L \frac{z^{-s}\, dz}{1-e^{-2\pi i z}} = 2\pi i \text{ (sum of residues of } \frac{z^{-s}}{1-e^{-2\pi i z}} \text{ at integers other than 0)}$$

$$= \sum_{m=1}^{\infty} (m^{-s} + e^{i\pi s} m^{-s}) = \zeta(s)(1+e^{i\pi s}).$$

This gives the analytic continuation of $\zeta(s)$ over the whole plane except possibly for even integers. Now, by (1.1),

$$\int_L u^{-s}\left(\frac{1}{1-e^{-2\pi i u}} - \frac{e^{-i\pi u^2}}{e^{i\pi u} - e^{-i\pi u}}\right) du = \int_L u^{-s}\left[\int_{0\nearrow 1} \frac{\exp(i\pi z^2 + 2\pi i u z)\, dz}{e^{i\pi z} - e^{-i\pi z}}\right] du$$

$$= \int_{0\nearrow 1} \frac{e^{i\pi z^2}}{e^{i\pi z} - e^{-i\pi z}} \left(\int_L u^{-s}\, e^{2\pi i u z}\, du\right) dz$$

$$= -2(2\pi)^{s-1} \sin s\pi\, \Gamma(1-s)\, e^{\frac{1}{2}is\pi} \int_{0\nearrow 1} \frac{e^{i\pi z^2}\, z^{s-1}}{e^{i\pi z} - e^{-i\pi z}}\, dz,$$

i.e.

$$\zeta(s)(1+e^{i\pi s}) = \int_L \frac{e^{-i\pi u^2} u^{-s}\, du}{e^{i\pi u} - e^{-i\pi u}} - 2(2\pi)^{s-1} \sin s\pi\, \Gamma(1-s)\, e^{\frac{1}{2}is\pi} \int_{0\nearrow 1} \frac{e^{i\pi z^2} z^{s-1}\, dz}{e^{i\pi z} - e^{-i\pi z}}.$$

In the first integral we may replace L by the two lines $0 \downarrow 1$ and $-1 \uparrow 0$, and we get

$$\int_L \frac{e^{-i\pi u^2} u^{-s}\, du}{e^{i\pi u} - e^{-i\pi u}} = \int_{0\downarrow 1} \frac{e^{-i\pi u^2} u^{-s}\, du}{e^{i\pi u} - e^{-i\pi u}} + \int_{-1\uparrow 0} \frac{e^{-i\pi u^2} u^{-s}\, du}{e^{i\pi u} - e^{-i\pi u}}$$

$$= \int_{0\downarrow 1} \left(\frac{e^{-i\pi u^2} u^{-s}}{e^{i\pi u} - e^{-i\pi u}} - \frac{e^{i\pi u^2} u^{-s} e^{i\pi s}}{e^{-i\pi u} - e^{i\pi u}} \right) du.$$

The curve may now be replaced by $0 \nwarrow 1$ if the sign is changed, and we thus obtain

$$-\int_{0\nwarrow 1} (1+e^{i\pi s}) \frac{e^{-i\pi u^2} u^{-s}\, du}{e^{i\pi u} - e^{-i\pi u}}.$$

The zeta-function is then expressed in the form

$$\zeta(s) = -\int_{0\nwarrow 1} \frac{e^{-i\pi u^2} u^{-s}\, du}{e^{i\pi u} - e^{-i\pi u}} - 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\, \Gamma(1-s) \int_{0\nearrow 1} \frac{e^{i\pi z^2} z^{s-1}\, dz}{e^{i\pi z} - e^{-i\pi z}} \qquad (2.1)$$

and the calculation of $\zeta(s)$ is reduced to that of the integral

$$I(s) = \int_{0\nearrow 1} \frac{e^{i\pi z^2} z^{-s}\, dz}{e^{i\pi z} - e^{-i\pi z}} = \int_{0\nearrow 1} h(z)\, dz; \qquad (2.2)$$

for

$$\overline{I(s)} = -\int_{0\nwarrow 1} \frac{e^{-i\pi z^2} z^{-\bar{s}}\, dz}{e^{-i\pi z} - e^{i\pi z}}.$$

If we multiply both sides of (2.1) by $\Gamma(\frac{1}{2}s)\pi^{-\frac{1}{2}s}$, and make use of the relation

$$\Gamma(1-s) = \frac{2^{-s}\pi^{\frac{1}{2}}\, \Gamma(\frac{1}{2}-\frac{1}{2}s)}{\sin \frac{1}{2}s\pi\, \Gamma(\frac{1}{2}s)},$$

we obtain

$$\zeta(s)\, \Gamma(\tfrac{1}{2}s)\, \pi^{-\frac{1}{2}s} = \Gamma(\tfrac{1}{2}s)\, \pi^{-\frac{1}{2}s}\, \overline{I(\bar{s})} - \Gamma(\tfrac{1}{2}-\tfrac{1}{2}s)\, \pi^{\frac{1}{2}s-\frac{1}{2}}\, I(1-s), \qquad (2.3)$$

and on the critical line this is equal to

$$-2i \sin \tfrac{1}{2}s\pi\, \Gamma(\tfrac{1}{2}s)\, \pi^{-\frac{1}{2}s}\, I(\bar{s}). \qquad (2.4)$$

For points on this line there is, therefore, only one real integral to calculate. For points not on the line there are four real integrals.

3. *The method of calculation.* Let μ be a positive real number lying between the integers m and $m+1$. Then

$$I(s) = \int_{0\nearrow 1} h(z)\, dz = \int_{m\nearrow m+1} h(z)\, dz - \sum_{r=1}^{m} r^{-s}.$$

Now let κ be a positive real number and put

$$g(z) = \frac{h(z)}{1-\exp[-2\pi\kappa\epsilon(z-\mu)]}.$$

The function g has simple poles at the integer points other than 0, and at the points $p_k = \mu + \epsilon k/\kappa$, where k is an integer; otherwise it is regular except on the closed positive imaginary axis. The residue at the non-zero integer r is

$$\frac{r^{-s}}{2\pi i\{1-\exp[-2\pi\kappa\epsilon(r-\mu)]\}},$$

and at p_k it is

$$\frac{h(p_k)}{2\pi\kappa\epsilon}.$$

The line on which the poles p_k lie, taken as running from left to right, will be called P.

Let J and J' be two curves going from the third quadrant to the first and from the first to the third respectively, J being entirely on the right of P and J' entirely on its left and on the right of the origin. Suppose also that there is a positive real number α such that, at sufficiently great distance from the origin, the curves lie in the region where either $\alpha < \arg z < \frac{1}{2}\pi - \alpha$ or $\frac{1}{2}\pi - \alpha > \arg z > -\pi + \alpha$, and that the length of curve with $|z| < R$ is $O(R)$. Then it is easily seen that

$$\int_{J+J'} g(z)\,dz = 2\pi i \text{ (sum of residues of } g \text{ between } J \text{ and } J')$$

$$= \Sigma \frac{r^{-s}}{1-\exp[-2\pi\kappa\epsilon(r-\mu)]} + \sum_k \frac{\epsilon}{\kappa} h(p_k),$$

where the first sum is taken over the integers r lying between J and J'. Now

$$\int_J g(z)\,dz = \int_J \big(g(z)-h(z)\big)\,dz + \int_J h(z)\,dz$$

$$= \int_J h(z)\,\frac{\exp[-2\pi\kappa\epsilon(z-\mu)]}{1-\exp[-2\pi\kappa\epsilon(z-\mu)]}\,dz + \int_P h(z)\,dz,$$

and

$$\int_{J'} g(z)\,dz = -\int_{J'} h(z)\cdot\frac{\exp[2\pi\kappa\epsilon(z-\mu)]}{1-\exp[2\pi\kappa\epsilon(z-\mu)]}\,dz.$$

If the curves J and J' are always distant more than $\frac{1}{4}\kappa^{-1}$ from the line P, then we have on them

$$\left|\frac{1}{1-\exp[2\pi\kappa\epsilon\,\mathrm{sg}(z)(z-\mu)]}\right| < 1{\cdot}27,$$

where $\mathrm{sg}(z)$ has the value 1 or -1 according as z is to the left or to the right of P. We can now collect our results in the form

$$I(s) = \sum_k \frac{\epsilon}{\kappa} h(p_k) - \sum_{r=1}^{\infty} r^{-s}\theta_r + R_0,$$

where θ_r has the value 1 if r is on the left of J', the value

$$\{1-\exp[2\pi\kappa\epsilon(r-\mu)]\}^{-1}$$

if r is between J and J', and the value 0 otherwise. The remainder R_0 satisfies

$$|R_0| < \int_{J+J'} 1{\cdot}27 \exp[-\sqrt{2\pi\kappa\mu}\,\mathrm{sg}(z)] \frac{|z|^{-\sigma}}{|e^{i\pi z}-e^{-i\pi z}|} e^{\Re\phi(z)}|dz|, \quad (3.2)$$

where

$$\phi(z) = i\pi z^2 + 2\pi\kappa\epsilon z\,\mathrm{sg}(z) - it\log z. \quad (3.3)$$

We may also write the formula for $I(s)$ in the form

$$I(s) = \sum_{k=-\infty}^{\infty} \frac{\epsilon}{\kappa} \frac{e^{i\pi p_k^2} p_k^{-s}}{e^{i\pi p_k}-e^{-i\pi p_k}} - \sum_{r=1}^{\infty} \frac{r^{-s}}{1-\exp[2\pi\kappa\epsilon(r-\mu)]} + R, \quad (3.4)$$

where $R = R_0 + R_1$, $p_k = \mu + \epsilon k/\kappa$ and

$$|R_1| < 1{\cdot}27\,\Sigma r^{-s} \exp[-\sqrt{2}\,\pi\kappa|r-\mu|],$$

the summation being over positive integers not between J and J'. By expressing $I(s)$ in this form we can eliminate from the numerical calculation any reference to the position of the curves J, J', and the remainder is not appreciably increased. In § 5 we choose the curves so that $R_1 = 0$. Of course, in the calculation the factor $\{1-\exp[2\pi\kappa\epsilon(r-\mu)]\}^{-1}$ will be put equal to 1 or to 0 except for a comparatively small number of terms.

In estimating the remainder we suppose that $\sigma \geqslant 0$, but this is not necessary.

4. *General remarks on the estimation of the remainder.* Suppose that

$$U = \int_{C_0} e^{w(z)} k(z)\,dz.$$

Then, for any curve C deformable into C_0 in the domain of regularity of $e^{w(z)}k(z)$,

$$|U| \leqslant \int_C e^{\Re w(z)} |k(z)|\,dz. \quad (4.1)$$

Now suppose that $\Re w(z)$ has large and fast variations, while $|k(z)|$ is comparatively steady. Then the value of the integral (4.1) is

principally affected by the maximum value of $\Re w(z)$ on C, and a good inequality for $|U|$ is obtained by minimising this maximum. It is easily seen that if there is a curve for which the maximum is minimised, and if z_0 is a point at which the maximum on this curve is attained, then $w'(z_0) = 0$, *i.e.* z_0 is a "saddle point" of w. Suppose that in the neighbourhood of the saddle point the curve is $z = z_0 + le^{i\alpha}$, l being an arc-length parameter. Then the contribution to the integral from the neighbourhood of the saddle point is approximately

$$e^{\Re w(z_0)} |k(z_0)| \int_{-\infty}^{\infty} |\exp[\tfrac{1}{2}w''(z_0)\, e^{2i\alpha} l^2]|\, dl = \frac{(2\pi)^{\frac{1}{2}} |k(z_0)|\, e^{\Re w(z_0)}}{\sqrt{\{-\Re[w''(z_0)\, e^{2i\alpha}]\}}}.$$

We naturally choose α to be $\frac{1}{2}\pi - \frac{1}{2}\arg w''(z_0)$, and then the expression becomes

$$\sqrt{\left\{\frac{2\pi}{|w''(z_0)|}\right\}}\, |k(z_0)\, e^{w(z_0)}|.$$

In the estimation of R_0 we could take $w(z)$ to be either $\phi(z)$ or $\phi(z) \pm i\pi z$, in the latter case different signs being taken on the two curves J, J'. With the first of these forms the analysis is simpler, but the second gives a better result; we deal only with the simpler form.

We actually use the idea of saddle-point integration in the following form. Suppose that we have a curve C with arc-length parameter l beginning at $l = 0$. Then

(*a*) if $\Re\psi'(z)\dfrac{dz}{dl} \leqslant -al$ on the curve, where $a > 0$, then

$$e^{\Re\psi[z(l)]} \leqslant e^{-\frac{1}{2}al^2} e^{\Re\psi[z(0)]}$$

on the curve, and consequently

$$\int_C e^{\Re\psi[z(l)]}\, |dz| \leqslant \sqrt{\left(\frac{\pi}{2a}\right)}\, e^{\Re\psi[z(0)]}.$$

This enables us to estimate the contribution to the integral from the neighbourhood of the saddle-point; we may estimate the contribution from the remainder of the curve by one or more applications of

(*b*) if $\Re\psi'(z)\dfrac{dz}{dl} \leqslant -a < 0$ on the curve, then $e^{\Re\psi[z(l)]} \leqslant e^{-al} e^{\Re\psi[z(0)]}$ and

$$\int_C e^{\Re\psi[z(l)]}\, |dz| \leqslant \frac{1}{a}\, e^{\Re\psi[z(0)]}.$$

5. *Detailed estimation of the remainder.* For the two curves J, J' we have two different saddle-points z_0, z_0', these being zeros of $\phi'(z)$ on the

two sides of the line P. We may put

$$\phi'(z) = 2\pi iz + 2\pi\kappa\epsilon - it/z = \frac{2\pi i}{z}(z-z_0)(z-z_1)$$

on the left of P, and

$$\phi'(z) = 2\pi iz - 2\pi\kappa\epsilon - it/z = \frac{2\pi i}{z}(z-z_0')(z-z_1')$$

on the right of P; the points z_0, z_0' are in the right half-plane and z_1, z_1' in the left. If we put further $\tau = t/2\pi$, $\rho = \kappa\tau^{-\frac{1}{2}}$, $\zeta = z\tau^{-\frac{1}{2}}$, we may write these equations as

$$(\zeta-\zeta_0)(\zeta-\zeta_1) = \zeta^2 + \bar{\epsilon}\rho\zeta - 1, \quad (\zeta-\zeta_0')(\zeta-\zeta_1') = \zeta^2 - \bar{\epsilon}\rho\zeta - 1.$$

The roots satisfy

$$\zeta_0+\zeta_1' = 0, \quad \zeta_1+\zeta_0' = 0, \quad \arg\zeta_0 + \arg\zeta_0' = 0, \quad \tfrac{1}{4}\pi > \arg\zeta_0' > 0.$$

There is a cubic curve independent of ρ on which all four roots lie. We shall need a number of other properties of the roots, and we mention them as we require them; they are mostly inequalities which can be proved by straightforward but laborious methods.

The behaviour of a number of functions of ρ is shown in Fig. 1.

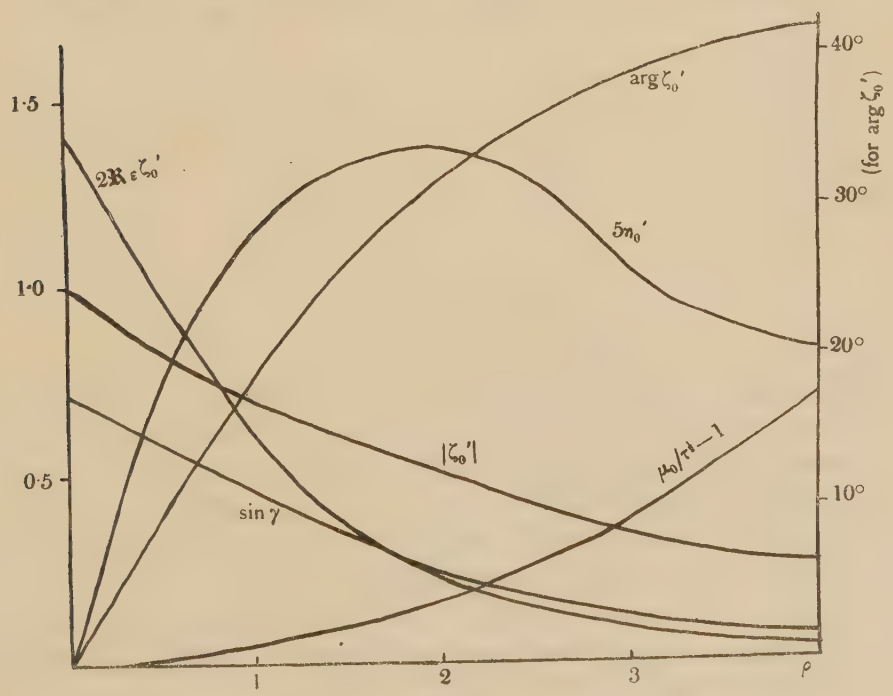


Fig. 1.

We choose the curves of integration J, J' as follows. J consists of three straight parts J_1, J_2, J_3, of which J_1 is a straight line from $-\epsilon\infty$ through z_0 to b, where $b = z_0 - \frac{1}{2}y_0(1+i)$; J_2 joins b to $b+\beta$, where β is real and positive; and J_3 joins $b+\beta$ to $\epsilon\infty$, passing through a half odd integer. As β tends to infinity the contribution to the remainder from J_3 tends to zero and the contribution from J_2 tends to a limit. We may therefore omit J_3 and suppose J to consist of J_1 together with J_2, taken as extending to infinity; there are then no poles on the right of J to contribute to R_1.

Fig. 2.

The curve J' consists of the four parts J_1', J_2', J_3', J_4', of which J_2' is further divided into J_5' and J_6'. Let $b' = z_0' - \frac{1}{2}y_0'(1+i)$. Then J_1' is a straight line from $\epsilon\infty$ to b' through z_0'; J_2' is a straight line from b' to c, where c/b' is real, $0 < c/b' < 1$, and $|c| = \min(\frac{1}{2}, \frac{1}{3}|z_0'|)$. The curve J_3' is part of a circle lying in the half-plane $\Re\epsilon z > 0$ and joining c to $-i\bar{c}$. We obtain J_4' by reflecting J_1' and J_2' in the line $\Im\epsilon z = 0$ and reversing its direction. We divide J_2' into J_5' and J_6', of which J_5' is the part for which $y > \frac{1}{3}y_0'$. Both of these parts are of positive length, J_5' being of length $\frac{1}{6}|z_0'|$ at least.

On J_1 and J_2 we have $|z| > \tau^{\frac{1}{2}} \mathfrak{R}\epsilon\zeta$ and

$$\left|\frac{1}{e^{i\pi z}-e^{-i\pi z}}\right| \leqslant \tfrac{1}{2}\operatorname{cosech}|\tfrac{1}{2}\pi y_0|.$$

The contribution to R_0 from J is therefore at most

$$\tfrac{1}{2}(1{\cdot}27)\tau^{-\frac{1}{2}\sigma}(\mathfrak{R}\epsilon\zeta_0)^{-\sigma}\exp[\sqrt{2\pi\kappa\mu}]\operatorname{cosech}|\tfrac{1}{2}\pi y_0|\int_J e^{\mathfrak{R}\phi(z)}|dz|. \quad (5.1)$$

On J_1 we have $0 > \arg(z-z_1)/z > -\cos^{-1}\frac{1}{3}$, a simple consequence of the facts that the distance of J_1 from 0 is greater than $2^{-\frac{1}{2}}|z_1|$ and that $\arg z_1 < -\frac{3}{4}\pi$. Also

$$|2\pi(z-z_1)(z-z_0)/z| > \pi|z-z_0|,$$

so that

$$\mathfrak{R}\frac{2\pi i(z-z_1)(z-z_0)}{z}\frac{dz}{dl} < -\tfrac{1}{3}\pi l = -\tfrac{1}{3}\pi|z-z_0|,$$

if l always decreases as z_0 is approached. Then, applying (a) of § 4, we have

$$\int_{J_1} e^{\mathfrak{R}\phi(z)}|dz| < \sqrt{6}\, e^{\mathfrak{R}\phi(z_0)}. \quad (5.2)$$

Also
$$\mathfrak{R}\phi(b) < \mathfrak{R}\phi(z_0) - \tfrac{1}{6}\pi y_0^2. \quad (5.3)$$

On J_2, $dz/dl = 1$ and

$$\mathfrak{R}\phi'(z) = \mathfrak{R}\left(2\pi i z - 2\pi\kappa\epsilon - \frac{2\pi i\tau}{z}\right) = -\pi y_0 - \sqrt{2}\,\pi\kappa - \frac{\pi\tau y_0}{x^2+\frac{1}{4}y_0^2}.$$

But $\tau < x_0^2$, since $\xi_0 > 1$ for all ρ, and therefore

$$\mathfrak{R}\phi'(z)\frac{dz}{dl} < \pi(-\sqrt{2}\,\kappa - 2y_0) = -2\pi y_0'. \quad (5.4)$$

Consequently, by (b) of § 4, (5.3) and (5.4),

$$\int_{J_2} e^{\mathfrak{R}\phi(z)}|dz| < \frac{e^{\mathfrak{R}\phi(z)}}{2\pi y_0'} < \frac{\exp[\mathfrak{R}\phi(z_0) - \frac{1}{6}\pi y_0^2]}{2\pi y_0'}. \quad (5.5)$$

Now let us turn to J'. If z is a point of $J_1' + J_2'$, then $-i\bar{z}$ is the corresponding point of J_4', and

$$\mathfrak{R}\phi(-i\bar{z}) = \mathfrak{R}[i\pi(-i\bar{z})^2 + 2\kappa\epsilon(-i\bar{z}) - it\log z - it\log(-i\bar{z}/z)]$$

$$\leqslant \mathfrak{R}[-i\pi(iz)^2 + 2\kappa\bar{\epsilon}iz - it\log z - \tfrac{1}{2}t\pi],$$

i.e.
$$\mathfrak{R}\phi(-i\bar{z}) \leqslant \mathfrak{R}\phi(z) - \tfrac{1}{2}t\pi.$$

Also $|-i\bar{z}| = |z|$ and

$$|\operatorname{cosec}(-i\pi z)| < |\operatorname{cosec}\pi z|,$$

as may be proved by the use of the product representation of the sine function, remembering that $|\arg z| < \frac{1}{4}\pi$. Consequently

$$\int_{J'} \frac{|z|^{-\sigma}}{|e^{i\pi z}-e^{-i\pi z}|} e^{\Re\phi(z)}|dz| < (1+e^{-\frac{1}{2}l\pi}) \int_{J_1'+J_2'+J_3'} \frac{|z|^{-\sigma}}{|e^{i\pi z}-e^{-i\pi z}|} e^{\Re\phi(z)}|dz|. \quad (5.6)$$

On J_1' we have $0 > \arg(z-z_1')/z > \gamma-\frac{1}{2}\pi$, where $\gamma = \frac{1}{4}\pi - \arg\zeta_0'$, and

$$|2\pi(z-z_1')(z-z_0')/z| > 2\pi|z-z_0'|.$$

Therefore
$$\Re\phi'(z)\frac{dz}{dl} < -2\pi\sin\gamma$$

on J_1', if l decreases as z_0' is approached from either side. Applying (a) of § 4, we have

$$\int_{J_1'} e^{\Re\phi(z)}|dz| < \frac{e^{\Re\phi(z_0')}}{(\sin\gamma)^{\frac{1}{2}}} \quad (5.7)$$

and
$$\Re\phi(b) < \Re\phi(z_0') - \tfrac{1}{2}\pi y_0'^2 \sin\gamma. \quad (5.8)$$

On J_2'

$$\begin{aligned}\Re\phi'(z)\frac{dz}{dl} &= \Re\frac{-2\pi iz}{|z|}\left(z+\bar{\epsilon}\kappa-\frac{\tau}{z}\right)\\ &\leqslant \Re\frac{-2\pi ib'}{|b'|}\left(b'+\bar{\epsilon}\kappa-\frac{\tau}{b'}\right)\\ &= \Re\frac{-2\pi i}{|b'|}(b'-z_0')(b'-z_1'). \end{aligned} \quad (5.9)$$

But $\arg(z_0'-b') = \frac{1}{4}\pi$ and

$$0 > \arg(b'-z_1') > \arg(-z_1') = \gamma-\tfrac{1}{4}\pi,$$

so that

$$\tfrac{1}{2}\pi+\gamma < \arg\frac{-2\pi i}{|b'|}(b'-z_0')(b'-z_1') < \tfrac{3}{4}\pi.$$

Also
$$\left|-2\pi i(b'-z_0')(b'-z_1')/|b'|\right| > \sqrt{2}\,\pi y_0',$$

so that

$$\Re\frac{-2\pi i}{|b'|}(b'-z_0')(b'-z_1') < -\sqrt{2}\,\pi y_0'\sin\gamma, \quad (5.10)$$

and applying (b) of § 4, and using (5.9), (5.10) and (5.8), we obtain

$$\int_{J_2'} e^{\Re\phi(z)}|dz| < (\sqrt{2}\,\pi y_0'\sin\gamma)^{-1}\exp[\Re\phi(z_0')-\tfrac{1}{2}\pi y_0'^2\sin\gamma]. \quad (5.11)$$

Also

$$\Re\phi(\tfrac{2}{3}b') < \Re\phi(z_0') - \tfrac{1}{2}\pi y_0'^2 \sin\gamma - \tfrac{1}{6}\sqrt{2}\,\pi y_0' |z_0'| \sin\gamma. \qquad (5.12)$$

On J_3'

$$\left| \arg \left\{ \frac{-2\pi i}{z} (z-z_0')(z-z_1') \frac{dz}{dl} \right\} \right| < 2 \sin^{-1} \tfrac{1}{3} < \tfrac{1}{4}\pi < \tfrac{1}{2}\pi - \gamma;$$

so that on both J_6' and J_3'

$$\Re \frac{2\pi i}{z} (z-z_0')(z-z_1') \frac{dz}{dl} < -\sqrt{2}\,\pi y_0' \sin\gamma.$$

This gives, using also (5.12),

$$\int_{J_6'+J_3'} e^{\Re\phi(z)} |dz|$$

$$< (\sqrt{2}\,\pi y_0' \sin\gamma)^{-1} \exp\left[\Re\phi(z_0') - \tfrac{1}{2}\pi y_0'^2 \sin\gamma - \tfrac{1}{6}\sqrt{2}\,\pi y_0' |z_0'| \sin\gamma\right]. \quad (5.13)$$

On J_1' we have

$$\left| \frac{z^{-\sigma}}{e^{i\pi z} - e^{-i\pi z}} \right| < \tfrac{1}{2}\tau^{-\frac{1}{2}\sigma} |\tfrac{1}{2}\zeta_0'|^{-\sigma} \operatorname{cosech} |\tfrac{1}{2}\pi y_0'|; \qquad (5.15)$$

on J_5'

$$\left| \frac{z^{-\sigma}}{e^{i\pi z} - e^{-i\pi z}} \right| < \tfrac{1}{2}\tau^{-\frac{1}{2}\sigma} |\tfrac{1}{3}\zeta_0'|^{-\sigma} \operatorname{cosech} |\tfrac{1}{3}\pi y_0'|; \qquad (5.16)$$

and on J_6', J_3'

$$\left| \frac{z^{-\sigma}}{e^{i\pi z} - e^{-i\pi z}} \right| < \frac{1}{2\pi} [\min(\tfrac{1}{2}, \tfrac{1}{3}\tau^{\frac{1}{2}} |\zeta_0'|)]^{-\sigma-1} \operatorname{cosec} \arg \zeta_0'. \qquad (5.17)$$

We may now collect our results to give an inequality for $|R|$. We use (5.1), (5.2), (5.5), (5.6), (5.7), (5.11), (5.13), (5.15), (5.16) and (5.17), and we make the exponents $\Re\phi(z_0) + \sqrt{2}\,\pi\kappa\mu$ and $\Re\phi(z_0') - \sqrt{2}\,\pi\kappa\mu$ more explicit by the use of the relation $z_0^2 = \bar{\epsilon}\kappa z_0 + \tau$:

$$|R| = |R_0| < 0.635 (\Re\epsilon\zeta_0)^{-\sigma} \tau^{-\frac{1}{2}\sigma} \operatorname{cosech} |\tfrac{1}{2}\pi y_0| \left\{ 2{\cdot}45 + \frac{0{\cdot}160}{y_0'} e^{-\frac{1}{6}\pi y_0'^2} \right\}$$

$$\times \exp\left[-\pi\kappa \Re\epsilon(z_0 - 2\mu) + 2\pi\tau \arg\zeta_0\right] + 0{\cdot}635(1 + e^{-\frac{1}{2}\pi t}) \tau^{-\frac{1}{2}\sigma} |\zeta_0'|^{-\sigma}$$

$$\times \left\{ (\sin\gamma)^{-\frac{1}{2}} 2^{\sigma} \operatorname{cosech} |\tfrac{1}{2}\pi y_0'| + (\sqrt{2}\,\pi y_0' \sin\gamma)^{-1} \exp\left[-\tfrac{1}{2}\pi y_0'^2 \sin\gamma\right] \right.$$

$$\times \left[3^{\sigma} \operatorname{cosech} |\tfrac{1}{3}\pi y_0'| + \frac{1}{\pi} |\zeta_0'|^{\sigma} \tau^{\frac{1}{2}\sigma} [\min(\tfrac{1}{2}, \tfrac{1}{3}\tau^{\frac{1}{2}} |\zeta_0'|)]^{-\sigma-1} \operatorname{cosec} \arg\zeta_0' \right.$$

$$\left. \left. \times \exp\left[-\tfrac{1}{6}\sqrt{2}\,\pi y_0' |z_0'| \sin\gamma\right] \right] \right\} \exp\left[-\pi\kappa \Re\epsilon(2\mu - z_0') + 2\pi\tau \arg\zeta_0'\right]. \quad (5.18)$$

At this point it may be as well to repeat the definitions of the various quantities appearing on the right-hand side of this inequality, in terms of s, κ:

$$\sigma = \Re s, \quad t = \Im s, \quad \tau = t/2\pi, \quad \rho = \kappa\tau^{-\frac{1}{2}}, \quad \epsilon = e^{\frac{1}{8}i\pi}.$$

The complex numbers ζ_0', ζ_1', ζ_1, ζ_0 are the roots of the equation

$$(\zeta^2-1)^2 = -i\rho^2\zeta^2$$

lying respectively in the first, second, third and fourth quadrants, and

$$\zeta = \xi+i\eta, \quad z = x+iy = \tau^{\frac{1}{2}}\zeta, \quad \gamma = \tfrac{1}{4}\pi - \arg\zeta_0',$$

where ξ, η, x, y are real.

The estimate (5.18) for R is somewhat complicated. It may be simplified considerably when ρ is not large. I give an estimate for the case $\rho \leqslant \frac{1}{2}$; we then have

$$\left.\begin{array}{ll} \sqrt{2}\,\Re\epsilon\zeta_0' > 1, & \Re\epsilon(1-\zeta_0') > 0{\cdot}45\rho, \\ -\eta_0 > \rho/2\sqrt{2}, & \sin\gamma > 0{\cdot}55, \\ \eta_0' > 0{\cdot}29\rho, & |\zeta_0'| > 0{\cdot}81. \end{array}\right\} \tag{5.19}$$

The result is that, for $\rho \leqslant \frac{1}{2}$, $t \geqslant 25$,

$$\begin{aligned} |R| = |R_0| < \tau^{-\frac{1}{2}\sigma}\Big[& 0{\cdot}76\,.\,2^{\frac{1}{2}-\frac{1}{2}\sigma}\operatorname{cosech} 0{\cdot}78(\kappa/\sqrt{2}) \\ & \times\Big\{2{\cdot}45+0{\cdot}40(\kappa/\sqrt{2})^{-1}\exp\Big(-0{\cdot}13(\kappa/\sqrt{2})^2\Big)\Big\}e^{-A}+0{\cdot}71(0{\cdot}81)^{\frac{1}{2}-\sigma} \\ & \times\Big[1{\cdot}91\,.\,2^{\sigma-\frac{1}{2}}\operatorname{cosech} 0{\cdot}64(\kappa/\sqrt{2})+1{\cdot}00(\kappa/\sqrt{2})^{-1}\exp\Big(-0{\cdot}14(\kappa/\sqrt{2})^2\Big) \\ & \times\Big\{1{\cdot}74\,.\,3^{\sigma-\frac{1}{2}}\operatorname{cosech} 0{\cdot}42(\kappa/\sqrt{2})+1{\cdot}7(1{\cdot}62)^{\sigma-\frac{1}{2}}\tau^{\frac{1}{2}\sigma+\frac{1}{2}} \\ & \qquad\times(\kappa/\sqrt{2})\exp\Big(-0{\cdot}13\tau^{\frac{1}{2}}(\kappa/\sqrt{2})\Big)\Big\}\Big]e^{-B}\Big], \end{aligned} \tag{5.20}$$

where

$$A = \pi\kappa\Re\epsilon(z_0-2\mu)+2\pi\tau\arg\zeta_0', \quad B = \pi\kappa\Re\epsilon(2\mu-z_0')-2\pi\tau\arg\zeta_0'.$$

In the case $\rho \geqslant \frac{1}{2}$ the estimation of the remainder from (5.18) may be made easier by the use of the following inequalities, which are valid for all positive ρ:

$$\Re\epsilon\zeta_0' > \frac{1}{\rho^3+\rho^2+\rho+\sqrt{2}}, \quad \eta_0' > \frac{0{\cdot}65\rho}{\rho^2+2}, \quad \sin\gamma > \frac{1}{\rho^2+\frac{1}{2}\rho+\sqrt{2}}. \tag{5.21}$$

6. *Choice of parameters. The remainder with a finite series.* The remainder R as given by (5.18) is the sum of two terms in which the main factors are e^{-A} and e^{-B}. The most favourable choice of the parameter μ is presumably approximately that which makes these two factors equal. Calling this value μ_0, we have

$$\mu_0 = \frac{1}{2\sqrt{2}}\left[\mathfrak{R}\epsilon(z_0'+z_0)+\frac{4\tau^{\frac{1}{2}}}{\rho}\arg\zeta_0'\right]$$

$$= \frac{\tau^{\frac{1}{2}}}{\sqrt{2}}\left[\mathfrak{R}\sqrt{\left(i+\tfrac{1}{4}\rho^2\right)}+\frac{2}{\rho}\arg\zeta_0'\right].$$

As ρ tends to infinity, $\mu_0 \sim \kappa/2\sqrt{2}$; and, as ρ tends to 0, $\mu_0 \sim \tau^{\frac{1}{2}}$. Also

$$\mathfrak{R}\epsilon(\mu_0-z_0') = \tfrac{1}{4}\left[\mathfrak{R}\epsilon(z_0-3z_0')+\frac{4\tau^{\frac{1}{2}}}{\rho}\arg\zeta_0'\right]$$

$$= \tfrac{1}{2}\kappa\left[1-\mathfrak{R}\sqrt{\left(\tfrac{1}{4}+\frac{i}{\rho^2}\right)}+\frac{2}{\rho^2}\arg\zeta_0'\right].$$

As ρ tends to 0, the factor

$$1-\mathfrak{R}\sqrt{\left(\tfrac{1}{4}+\frac{i}{\rho^2}\right)}+\frac{2}{\rho^2}\arg\zeta_0'$$

tends to 1; and, as ρ tends to infinity, this factor tends to $\frac{1}{2}$. For all positive values of ρ it is greater than $\frac{1}{2}$, and therefore

$$\mathfrak{R}\epsilon(\mu_0-z_0') > \tfrac{1}{4}\kappa.$$

Also
$$\mathfrak{R}\epsilon(z_0-\mu_0) = \tfrac{1}{4}\left[\mathfrak{R}\epsilon(3z_0-z_0')-\frac{4\tau^{\frac{1}{2}}}{\rho}\arg\zeta_0'\right],$$

$$= \tfrac{1}{2}\kappa\left[1+\mathfrak{R}\sqrt{\left(\tfrac{1}{4}+\frac{i}{\rho^2}\right)}-\frac{2}{\rho^2}\arg\zeta_0'\right]$$

and we have
$$\mathfrak{R}\epsilon(z_0-\mu_0) > \tfrac{1}{2}\kappa.$$

If for μ we choose $\mu_0 \pm \delta$, where $\delta > 0$, then the greater of the exponential factors e^{-A}, e^{-B} is

$$\exp\left[-\tfrac{1}{2}\pi\kappa^2+\sqrt{2}\,\pi\kappa\delta\right].$$

The values of μ which we may choose are restricted only by the condition that the curves J, J' must be distant at least $1/(4\kappa)$ from P. If $\kappa \geqslant \sqrt{2}$, we may then choose μ in the interval $(\mu_0, \mu_0+\frac{1}{2})$; and, if $\kappa \geqslant 2$, we may choose it in the interval $(\mu_0-\frac{1}{2}, \mu_0+\frac{1}{2})$. However, for such small values of κ it will probably be best to choose μ rather close to μ_0. We need not consider the case of smaller values of κ than $\sqrt{2}$, since, as will appear,

it is not advantageous to take such small values, even when there is only one term taken from the series $\Sigma h(p_k)$.

When ρ is small, μ_0 is close to $\tau^{\frac{1}{2}}$, and it is therefore probably simplest to choose a value of μ which is close to $\tau^{\frac{1}{2}}$ without actually calculating μ_0. The inequality

$$0 < \mu_0/\tau^{\frac{1}{2}} - 1 < \rho^2/3 \quad (0 < \rho < 1)$$

should then be of value.

For large values of κ (*e.g.* $\kappa > 3$) we shall do well to choose μ to be either an integer or a half odd integer. In the case that μ is an integer the function g has a double pole at μ; in place of the terms

$$\frac{\epsilon}{\kappa} h(\mu) - \frac{\tau^{-s}}{1-\exp[2\pi\kappa\epsilon(\mu-\mu)]}$$

we have therefore to put the residue at 0 of

$$\frac{2\pi i(-)^{\mu}(z+\mu)^{-s}\exp[i\pi(z+\mu)^2 - i\pi z]}{(1-e^{-2\pi i z})(1-e^{-2\pi\kappa\epsilon z})}$$

and this is equal to

$$\mu^{-s}\left\{-\tfrac{1}{2} + \frac{\sigma}{2\pi\kappa\epsilon\mu} - \frac{\epsilon}{\kappa}\left(\mu - \frac{t}{2\pi\mu}\right)\right\}.$$

In the practical applications of course we take only a finite number of terms from the series $\Sigma(\epsilon/\kappa)h(p_k)$. We therefore want an estimate of the error arising from this. If we decide what is the greatest total error η we can admit in our result we can proceed in this way. We choose κ so that $|R| < \frac{1}{2}\eta$ and then take sufficiently many terms of the series for the error from this second source not to exceed $\frac{1}{2}\eta$. Now let us estimate this second remainder. For this purpose we prove the following

LEMMA. *The function* $|e^{i\pi z^2} z^{-it}|$ *has only one maximum on the line* P.

Put $z = \mu\left(1+\theta(1+i)\right)$, $a = t/(2\pi\mu^2)$; then θ is real and

$$\log|e^{i\pi z^2} z^{-it}| = \Re\left[i\pi\left(1+\theta(1+i)\right)^2 - 2\pi i a \log\left\{\mu\left(1+\theta(1+i)\right)\right\}\right]\mu^2.$$

Let us abbreviate the right-hand side to $H(\theta)$. Then

$$H'(\theta) = 2\pi\mu^2\left(-1-2\theta+\frac{a}{(1+\theta)^2+\theta^2}\right) = -2\pi\mu^2\,\frac{\left((1+\theta)^2+\theta^2\right)(1+2\theta)-a}{(1+\theta)^2+\theta^2}.$$

Now

$$\frac{d}{d\theta}\left[\left((1+\theta)^2+\theta^2\right)(1+2\theta)-\alpha\right]=3(2\theta+1)^2+1>0,$$

and therefore $\left((1+\theta)^2+\theta^2\right)(1+2\theta)-\alpha$ cannot vanish for more than one value of θ; it clearly vanishes for at least one value. Then $H(\theta)$ has just one stationary value, which is easily seen to be a maximum. This completes the proof of the lemma. If $\alpha<1$ the value θ giving the maximum satisfies $0>\theta>\alpha^{\frac{1}{3}}-1$.

Put $u_k=\kappa^{-1}|h(p_k)|$; then, if $\sigma\geqslant 0$,

$$\sum_{k=K+1}^{\infty} u_{-k}<\frac{(\mu/\sqrt{2})^{-\sigma}}{1-\exp\left[\sqrt{2}\pi(K+1)/\kappa\right]}\sum_{k=K+1}^{\infty}\exp\left(-\frac{\pi k}{\sqrt{2}\kappa}\right)|e^{i\pi p_{-k}^2}p_{-k}^{-it}|.$$

If we suppose that $\Im(p_{-K-1})$ is less than the value of y for which the maximum of $|e^{i\pi z^2}z^{-it}|$ occurs, and if $k\geqslant K+1$, then

$$|e^{i\pi p_{-k}^2}p_{-k}^{-it}|\leqslant|e^{i\pi p_{-K-1}^2}p_{-K-1}^{-it}|$$

and therefore

$$\sum_{K+1}^{\infty} u_{-k}<\left(\frac{\mu}{\sqrt{2}}\right)^{-\sigma}\frac{\exp\left[-\pi(K+1)/\kappa\sqrt{2}+\Re(i\pi p_{-K-1}^2)\right]|p_{-K-1}^{-it}|}{\left(1-\exp\left[-\sqrt{2}\pi(K+1)/\kappa\right]\right)\left(1-\exp\left[-\pi/\kappa\sqrt{2}\right]\right)}$$

$$<\frac{2^{\frac{1}{2}\sigma}u_{-K-1}}{\left(1-\exp\left[-\sqrt{2}\pi(K+1)/\kappa\right]\right)^2\left(1-\exp\left[-\pi/\kappa\sqrt{2}\right]\right)}.$$

Similarly, if (as is always the case if $K'\geqslant 0$ and $\alpha\leqslant 1$) $\Im(p_{K'+1})$ is greater than the value of y for which the maximum occurs,

$$\sum_{K'+1}^{\infty} u_k<\frac{u_{K'+1}}{\left(1-\exp\left[-\sqrt{2}\pi(K'+1)/\kappa\right]\right)^2\left(1-\exp\left[-\pi/\kappa\sqrt{2}\right]\right)},$$

so that

$$|R^*|=\left|\left(\sum_{-\infty}^{\infty}-\sum_{-K}^{K'}\right)\frac{\epsilon}{\kappa}h(p_k)\right|$$

$$<\frac{2^{\frac{1}{2}\sigma}|h(p_{-K-1})|+|h(p_{K'+1})|}{\kappa\left(1-\exp\left[-\sqrt{2}\pi(K^*+1)/\kappa\right]\right)^2\left(1-\exp\left[-\pi/\kappa\sqrt{2}\right]\right)},$$

where $K^*=\min(K, K')$.

In the case in which κ is small compared with $\tau^{\frac{1}{3}}$ we can easily obtain a rough estimate of the number of terms required for a given accuracy. For in this case K, K' are such that $p_{-K}/\mu-1$ and $p_{K'}/\mu-1$ are small,

and u_K is approximately $\exp[-2\pi(K+1)^2/\kappa^2]$. If the remainders R and R^* are of the same order of magnitude, then we have approximately $2\pi(K+1)^2/\kappa^2 = \frac{1}{2}\pi\kappa^2$, *i.e.*, $K+1 = \frac{1}{2}\kappa^2$. The number of terms T that we take is $2K+1$, *i.e.*, approximately κ^2, and the total error is of the order of magnitude of $e^{-\frac{1}{2}\pi T}$. If this statement is to be put into an exact form we must say that if μ and κ are suitably chosen as functions of t, η, and σ lies in the interval $0 \leqslant \sigma \leqslant 1$, and if η tends to 0 and t to infinity in such a way that κ^{-1}, $\kappa\tau^{-\frac{1}{2}}$ also tend to 0, then the error does not exceed η and $T^{-1}\log\eta^{-1}$ tends to $\frac{1}{2}\pi$; this does not hold for any number larger than $\frac{1}{2}\pi$.

When $\kappa\tau^{-\frac{1}{2}}$ is of the order of magnitude of 1 we cannot get so simple an estimate of the number of terms needed, but we can obtain an estimate in the limiting case $\kappa\tau^{-\frac{1}{2}} \to \infty$. We may then neglect all the factors in u_h except $|e^{i\pi}p_k^2|$. Putting $\mu+\epsilon v = p_{-K}$, we have approximately

$$|e^{i\pi(\mu+\epsilon v)^2}| = e^{-\frac{1}{2}\pi\kappa^2}$$

if R and R^* are of the same order of magnitude; *i.e.*,

$$\sqrt{2}\pi\mu v + \pi v^2 = \tfrac{1}{2}\pi\kappa^2.$$

But for large values of μ we have approximately $\mu = \kappa/2\sqrt{2}$, and therefore approximately

$$2v^2 + \kappa v - \kappa^2 = 0.$$

The two roots of this equation are $\bar{\epsilon}(P_{-K}-\mu)$ and $\bar{\epsilon}(P_{K'}-\mu)$ (approximately); the difference of the roots is $\frac{3}{2}\kappa$ and therefore $T = \frac{3}{2}\kappa^2$ approximately, and the error is of the order of magnitude of $e^{-\frac{1}{3}\pi T}$.

It is possible that this might be improved by taking μ to be something different from μ_0; for if we take μ closer to $\tau^{\frac{1}{2}}$ the remainder R^* is made smaller for any given value of T. Such an improvement would necessarily be at the expense of the remainder R; I do not think that any appreciable improvement really can be made along these lines.

In the case in which $T = 3$ we may put $\kappa = 1{\cdot}6\sqrt{2}$, and then, if $\sigma = \frac{1}{2}$, $\mu = \tau^{\frac{1}{2}}$ and $t > 350$, and if the factor $\left(1-\exp[2\pi\kappa\epsilon(r-\mu)]\right)^{-1}$ is replaced by 0 or 1 except for two terms of the main series, the error from all sources does not exceed $0{\cdot}0044\tau^{-\frac{1}{4}}$.

7. *A similar method.* There is an alternative, and better known, integral representation of the zeta-function on which we may base our calculations, viz.,

$$\zeta(s) = \sum_{r=1}^{m} r^{-s} + 2(2\pi)^{s-1}\sin\tfrac{1}{2}\pi s\,\Gamma(1-s)\sum_{r=1}^{m'} r^{-s} + \frac{1}{1+e^{i\pi s}}\int_{Q_m}\frac{e^{2m'i\pi z}z^{-s}\,dz}{1-e^{-2\pi iz}}.$$

Here Q_m is a curve coming from infinity in the first quadrant, crossing the real axis between m and $m+1$ and again between $-m$ and $-m-1$, and going on to infinity in the second quadrant; z^{-s} is defined as in § 2. If we choose $m = m' = [\tau^{\frac{1}{2}}]$, and let the part of Q_m in the neighbourhood of the positive real axis be a straight line cutting the negatively directed real axis at μ and at an angle of $+\frac{1}{4}\pi$, then, for large t, the only appreciable contribution to the integral comes from the neighbourhood of the positive real axis. We can approximate to this integral in the same way as before, the resulting approximate value for $\zeta(s)$ being

$$\sum_{r=1}^{\infty} r^{-s}\left\{1-(1+e^{i\pi s})^{-1}\left(1-\exp\left[-2\pi\kappa\epsilon(r-\mu)\right]\right)^{-1}\right\}$$

$$+2(2\pi)^{s-1}\sin\tfrac{1}{2}\pi s\,\Gamma(1-s)\sum_{r=1}^{m} r^{-s}$$

$$-\frac{\epsilon}{\kappa(1+e^{i\pi s})}\sum_{k=-K}^{K}\frac{(\mu+\epsilon k/\kappa)^{-s}\exp\left[2\pi i m(\mu+\epsilon k/\kappa)\right]}{1-\exp\left[-2\pi i(\mu+\epsilon k/\kappa)\right]}.$$

The integer K must not be chosen too large; $K < \frac{1}{2}\tau^{\frac{1}{2}}\kappa$ is usually sufficiently small. This method has the advantage that for points not on the critical line only two real integrals have to be evaluated, and not four. This may be of value for calculation of zeros not on the critical line. For this purpose it will not matter that the method is only applicable for large values of t; it is, however, possible to remove this restriction by integrating along a parabola, *e.g.*, the parabola

$$x^2 = 2\mu y+\mu^2 \quad ([\tau^{\frac{1}{2}}] \leqslant \mu \leqslant [\tau^{\frac{1}{2}}]+1).$$

The conformal mapping $u^2 = z$ transforms this parabola into a straight line, so that

$$\int_{Q_m}\frac{e^{2\pi i m z}z^{-s}\,dz}{1-e^{-2\pi i z}} = -2\int_{m^{\frac{1}{2}}\nearrow(m+1)^{\frac{1}{2}}}\frac{e^{2\pi i m u^2}u^{-2s+1}\,du}{1-e^{-2\pi i u^2}}.$$

The line of integration cuts the imaginary axis between $-im^{\frac{1}{2}}$ and $-i(m+1)^{\frac{1}{2}}$.

King's College,
Cambridge.

ON NUMBERS REPRESENTED BY BINARY CUBIC FORMS

By L. J. MORDELL.

[Received 10 October, 1940.—Read 17 December, 1940.]

1. *Introduction.*

Let $$f(x, y) = ax^3+bx^2y+cxy^2+dy^3, \tag{1}$$

where a, b, c, d are real numbers, be a binary cubic form. An important question is to find for the minimum of $|f(x, y)|$, where x, y are integers, not both zero, an upper bound expressed in terms of the determinant or negative discriminant D, the well-known invariant of $f(x, y)$, defined by

$$D = 27a^2d^2-18abcd-b^2c^2+4ac^3+4db^3. \tag{2}$$

Some results have been known for more than eighty years from the work of Eisenstein, Arndt, Hermite and others. Thus, if $D \neq 0$, integers x, y, not both zero, exist such that

$$|f(x, y)| \leqslant (|D|/l)^{\frac{1}{4}}, \tag{3}$$

where l is a number independent of the coefficients of $f(x, y)$. This was proved* with $l = 45\frac{9}{16} = (6\frac{3}{4})^2$ when $D < 0$ by Arndt in 1858, and by Hermite in 1859; the latter also proved it with $l = 16$ when $D > 0$. It has apparently been overlooked that these are not the best possible results; so far as I am aware, no suggestion to this effect has been made by other writers.

Some light is thrown upon the subject by considering the classical results for binary quadratic forms. Let

$$F(x, y) = Ax^2+Bxy+Cy^2,$$

* The most easily accessible reference is Hermite, "Sur la réduction des formes cubiques à deux indéterminées", *Oeuvres* (2), (1908), 93–98, or *Comptes rendus*, 48 (1859), 1, 351. He was concerned primarily with the reduction of binary cubic forms.

where A, B, C are real numbers, be a binary quadratic form of discriminant

$$\Delta = B^2 - 4AC.$$

Then, when $\Delta < 0$, integers x, y, not both zero, exist such that

$$|F(x, y)| \leqslant (\tfrac{1}{3}|\Delta|)^{\frac{1}{2}},$$

the equality sign being necessary when, and only when, on denoting by E an arbitrary constant,

$$E^{-1}F(x, y) \sim x^2 + xy + y^2,$$

of discriminant -3; and when $\Delta > 0$, integers x, y, not both zero, exist such that

$$|F(x, y)| \leqslant (\tfrac{1}{5}\Delta)^{\frac{1}{2}},$$

the equality sign occurring when, and only when,

$$E^{-1}F(x, y) \sim x^2 + xy - y^2,$$

of discriminant 5.

The equivalence sign refers to unimodular substitutions, *i.e.*, those with integer coefficients and determinant ± 1.

These results also show that $x^2 + xy \pm y^2$ are the respective positive definite and indefinite rationally indecomposable binary quadratic forms with integer coefficients for which $|\Delta|$ assumes its least possible value. For, if $\Delta > -3$ or $\Delta < 5$ respectively, then $|F(x, y)| < 1$ for integer values of x, y, not both zero, and so $F(x, y) = 0$ for these values. This is impossible and so $|\Delta|$ is not less than 3 or 5 respectively.

The binary quadratic results, however, suggest the corresponding results for the binary cubic.

Suppose, first, that $D < 0$, so that $f(x, 1) = 0$ has three real roots. It is known from other considerations that, if a, b, c, d are integers and the cubic has no rational linear factors, then $D \leqslant -49$, the equality sign holding when, and only when,

$$f(x, y) \sim x^3 + x^2 y - 2xy^2 - y^3.$$

Suppose next that $D > 0$, so that $f(x, 1) = 0$ has only one real root. If, again, a, b, c, d are integers and the cubic has no rational linear factors, then $D \geqslant 23$, the equality sign holding when, and only when,

$$f(x, y) \sim x^3 - xy^2 - y^3.$$

2. *Formulation of the main theorems.*

These results suggest, but are not used in the proof of, the following theorems.

THEOREM 1. *If $D<0$, integer values of x, y, not both zero, exist such that*

$$|f(x, y)| \leqslant \sqrt[4]{\left(\frac{|D|}{49}\right)}, \tag{4}$$

where the equality sign is necessary when and only when $D=-49e^4$ and

$$e^{-1}f(x, y) \sim x^3+x^2y-2xy^2-y^3, \tag{5}$$

where e is an arbitrary constant.

THEOREM 2. *If $D>0$, integer values of x, y, not both zero, exist such that*

$$|f(x, y)| \leqslant \sqrt[4]{\left(\frac{D}{23}\right)}, \tag{6}$$

where the equality sign is necessary when and only when $D=23e^4$ and

$$e^{-1}f(x, y) \sim x^3-xy^2-y^3, \tag{7}$$

where e is an arbitrary constant.

For the sake of completeness we add and prove at once the comparatively trivial

THEOREM. *If $D=0$, then, for arbitrary $\epsilon>0$, integer values of x, y, not both zero, exist such that*

$$|f(x, y)|<\epsilon. \tag{8}$$

For now $f(x, y)$ can be written as

$$f(x, y)=a(x+py)^2(x+qy),$$

where a, p, q are all real.

Suppose first that $p \neq q$. Let N be an arbitrarily large number. Then by Minkowski's classical theorem (Lemma 5 of this paper), integer values of x, y, not both zero, exist such that

$$|x+py| \leqslant 1/N, \quad |x+qy| \leqslant N|p-q|.$$

Hence

$$|f(x, y)| \leqslant a|p-q|/N,$$

and the result follows.

The theorem is also obviously true when $p=q$.

3. *Alternative forms of the theorems.*

Theorems 1 and 2 will be stated and proved in a slightly different form.

Let a lattice $\mathfrak{L}$ of determinant unity be given by

$$x = \alpha\xi+\beta\eta, \quad y = \gamma\xi+\delta\eta, \tag{9}$$

where α, β, γ, δ are real numbers, $\alpha\delta-\beta\gamma = 1$, and ξ, η run through all integer values. Denote by O the origin $(x, y) = (0, 0)$.

THEOREM 3. *A point of $\mathfrak{L}$ other than O exists such that*

$$|g(x, y)| = |x^3+x^2y-2xy^2-y^3| \leqslant 1. \tag{10}$$

The equality sign is needed when and only when $\mathfrak{L}$ is equivalent either to the lattice $x = \xi$, $y = \eta$, or to one of three others defined by

$$(\theta-\phi)x = \theta\phi^2\xi+\psi\eta, \quad (\theta-\phi)y = -\theta^2\xi-\theta\phi^2\eta,$$

where θ, ϕ, ψ are the roots of $t^3-t^2-2t+1 = 0$.

THEOREM 4. *A point of $\mathfrak{L}$, other than O, exists such that*

$$|h(x, y)| = |x^3-xy^2-y^3| \leqslant 1. \tag{11}$$

The equality sign is needed when and only when $\mathfrak{L}$ is equivalent either to the lattice $x = \xi$, $y = \eta$, or to the one defined by

$$(3\theta^2-1)x = -\xi-(\theta+3)\eta, \quad (3\theta^2-1)y = -3\theta\xi+\eta,$$

where θ is the real root of $t^3-t-1 = 0$.

A number of alternative statements of the various results are given.

THEOREM 5. *Let p, q, r, p', q', r' be six numbers, which in case* (I) *are all real, while in case* (II) *p, p' are real and q, r are conjugate complex numbers, as are also q', r'. Suppose also that*

$$\Pi(qr'-q'r) \neq 0.$$

Let $K = \frac{1}{7}$, $1/\sqrt{(23)}$ in the respective cases. Then integer values of x, y, not both zero, exist such that

$$|(px+p'y)(qx+q'y)(rx+r'y)| \leqslant |K\Pi(q'r-qr')|^{\frac{1}{3}}. \tag{12}$$

Let $\mathfrak{L}$ be a lattice of determinant unity.

THEOREM 6. *A point of $\mathfrak{L}$ not O exists such that*

$$|xy(x+y)| \leqslant \sqrt{\tfrac{1}{7}}. \tag{13}$$

THEOREM 7. *A point of $\mathfrak{L}$ not O exists such that*

$$|y(x^2+y^2)| \leqslant \sqrt[4]{\tfrac{4}{23}}. \tag{14}$$

THEOREM 8. *A point of $\mathfrak{L}$ not O exists such that*

$$|x^3+y^3| \leqslant \sqrt[4]{\tfrac{27}{23}}. \tag{15}$$

These are all best possible results.

4. *The equivalence of Theorems* 1 *and* 3.

It is easy to see that any cubic $f(x, y)$, where $D < 0$, can be transformed by a real linear substitution into any other cubic $f'(x, y)$ with the determinant $D' < 0$. It suffices to show that if p, q, r, p', q', r' are all real, then

$$\Pi(q-r)(x-py) = \Pi'(q'-r')(x'-p'y')$$

for a suitable real linear substitution. Such a one is obviously given by solving for x, y in terms of x', y' from

$$(q-r)(x-py) = (q'-r')(x'-p'y'),$$

$$(r-p)(x-qy) = (r'-p')(x'-q'y'),$$

$$(p-q)(x-ry) = (p'-q')(x'-r'y');$$

these three equations reducing to two since their sum is zero

If we suppose in (9) that $\alpha\delta-\beta\gamma$ is not necessarily 1, and write now

$$f(\xi, \eta) = g(x, y),$$

the determinant D of $f(\xi, \eta)$ is $-49(\alpha\delta-\beta\gamma)^6$ from the invariant property of D, and so (4) becomes

$$|g(x, y)| \leqslant |\alpha\delta-\beta\gamma|^{\frac{3}{7}}.$$

This inequality is homogeneous in α, β, γ, δ and so there is no loss of generality if we suppose that $\alpha\delta-\beta\gamma = 1$, *i.e.* (4) follows from (10). Also (5) follows since from § 16, (10) is changed into itself by the linear substitution for x, y in terms of ξ, η given after (10).

Similarly, Theorems 2 and 4 are equivalent, since any cubic $f(x, y)$, where $D > 0$, can be transformed by a real linear substitution into any other cubic $f'(x, y)$ with its determinant $D' > 0$. For if, now, p, p' are real and q, q' are conjugate complex numbers as are also r, r', the preceding substitution still suffices. The rest of the argument is the same with $g(x, y)$, -49 replaced by $h(x, y)$, 23.

5. *Conjectured theorems*.*

Analogy with the binary quadratic form suggests the following extension of these results.

HYPOTHESIS 1. *If $D<0$ and $f(x, y)$ is not equivalent to*

$$x^3+x^2y-2xy^2-y^3,$$

then integer values of x, y, not both zero, exist such that

$$|f(x, y)| \leqslant \sqrt[4]{\left(\frac{|D|}{81}\right)},$$

where the equality sign holds when and only when $D=-81e^4$ and

$$e^{-1}f(x, y) \sim x^3-3xy^2+y^3,$$

of determinant -81.

HYPOTHESIS 2. *If $D>0$ and $f(x, y)$ is not equivalent to $x^3-xy^2-y^3$, then integer values of x, y, not both zero, exist such that*

$$|f(x, y)| \leqslant \sqrt[4]{\left(\frac{D}{31}\right)},$$

where the equality sign holds when and only when $D=31e^4$ and

$$e^{-1}f(x, y) \sim x^3+xy^2+y^3,$$

of determinant 31.

There are corresponding extensions for Theorems 3 and 4. The truth of these hypotheses may possibly be settled by the method of this paper.

6. *A related theorem.*

It may be noted that the theorems of this paper deal with the first lattice problems in two variables and of degree greater than two in which a simple non-convex area is considered and for which a best possible result has been demonstrated. It may be remarked, however, that Minkowski has considered a not altogether unrelated problem†. This is to find‡ the

---

* These hypotheses have been since disproved by Davenport, *Proc. Cambridge Phil. Soc.*, 37 (1941), 325–330. I retain them as the last word has not yet been said upon the subject.

† H. Minkowski, *Diophantische Approximation* (Leipzig, 1907), 51–58.

‡ *Note added* 23 *March*, 1943, in reading the proofs. I have since shown that the ideas developed in this paper can be extended to deal not only with the case $0<p<1$, but with more general non-convex regions. See *Journal London Math. Soc.*, 16 (1941), 149.

least possible value of λ_p, a function of p only, a given number greater than or equal to 1, such that the convex region defined by

$$|x|^p+|y|^p \leqslant \lambda_p{}^p$$

should contain a point other than O of every lattice $\mathfrak{L}$ of determinant unity. The problem is equivalent to that of finding the least positive value of Δ for which some lattice of determinant Δ has no inner point other than O in the region

$$|x|^p+|y|^p \leqslant 1.$$

Minkowski proves that the least value of Δ arises from lattices for which $(\xi, \eta) = (1, 0), (0, 1), (1, -1)$ give points on the boundary. He then had to find the minimum value of

$$\Delta = \alpha\delta-\beta\gamma,$$

where the parameters are subject to the conditions

$$|\alpha|^p+|\gamma|^p = 1, \quad |\beta|^p+|\delta|^p = 1, \quad |\alpha-\beta|^p+|\gamma-\delta|^p = 1.$$

He proposed a hypothesis for the minimum value which he stated geometrically, but I am not aware that this has been proved even for $p = 3$. Before I found the method of this paper, it seemed to me that Minkowski's method could be applied to the present problem. I worked out all the details on that assumption. The justification seemed to be difficult because an infinite region was involved. The present proof shows that actually it suffices to consider only a finite region, and in view of this it is possible that Minkowski's method might now apply. I hope to return to this later.

7. *Lattices.*

Some results on lattices of determinant unity are gathered here for convenience. The last is new and is fundamental in the proof of the main theorems. The others, except possibly Lemma 4, are well known and I do not prove them all.

Let $P(x_1, y_1)$, $Q(x_2, y_2)$ be two points of $\mathfrak{L}$, neither of which is O, the origin, and such that O, P, Q are not collinear.

LEMMA 1. *Twice the area of the triangle OPQ is a positive integer.*

LEMMA 2. *If the area of the triangle OPQ is $\frac{1}{2}$, the lattice can be generated by vector addition from OP, OQ.*

LEMMA 3. *If the area of the triangle OPQ is 1, and there are no points of $\mathfrak{L}$ on the lines OP, OQ except O, P, Q, then the mid-point M of PQ is a point of $\mathfrak{L}$.*

For let R be the remaining vertex of the parallelogram $OPRQ$. Let the line through M parallel to OP meet OQ in Q' and PR in R'. Then $\mathfrak{L}$ can be generated by vector addition from OP and OS where S is a point of $Q'R'$. For the area shows that S must lie on the line through Q', R', and then it is easily seen that S can be taken in the interval $Q'R'$. If S is not the mid-point of PQ, it lies either in POQ or in PRQ. In the first case, a contradiction arises from

$$1 = \text{area } POQ > \text{area } SOP + \text{area } SOQ = \tfrac{1}{2} + \tfrac{1}{2}.$$

Similarly if S lies in PRQ.

A region is said to contain a point when the point is an inner point or boundary point of the region.

LEMMA 4. *A parallelogram with one vertex at O and area unity cannot contain the points P, Q of $\mathfrak{L}$ unless one, P say, coincides with a vertex on a side of the parallelogram through O, and the other point Q lies on the side parallel to OP.*

For if this is not so, it is easily shown that the area of the triangle must be less than $\frac{1}{2}$, in contradiction to Lemma 1. For if the parallelogram is denoted by $OACB$, it suffices to take P in AC and Q in BC. Then

$$2 \text{ area } OPQ = 1 - \frac{AP}{AC}\frac{BQ}{BC} < 1,$$

unless $AP = 0$ or $BQ = 0$. Hence the result.

LEMMA 5. *Every parallelogram with area 4 and centre at O contains a point of $\mathfrak{L}$ other than O.*

This is Minkowski's classical theorem.

LEMMA 5a. *If the parallelogram in Lemma 5 contains no inner point of $\mathfrak{L}$ other than O, and one of its vertices is a point of $\mathfrak{L}$, then the middle points of its sides are also points of $\mathfrak{L}$.*

This is a very simple particular case of a result due to Minkowski.

LEMMA 6. *Let R and D be two given regions, not necessarily convex or finite, of which R contains the origin and is transformed into itself by a linear homogeneous substitution Σ, and D is such that no point of D is contained in R. Suppose that a lattice $\mathfrak{L}$ of determinant 1 exists such that no point of $\mathfrak{L}$ except O is contained in R and that every such lattice $\mathfrak{L}$ has a point contained in D. Then the region ΣD, i.e. the transform of D by Σ, also contains a point of $\mathfrak{L}$ other than O.*

For no point of the lattice $\Sigma^{-1}\mathfrak{L}$ except O can be contained in R, since then a point of $\Sigma\Sigma^{-1}\mathfrak{L}$ is contained in ΣR, *i.e.* a point of $\mathfrak{L}$ in R. Hence a point of $\Sigma^{-1}\mathfrak{L}$ not O is contained in D and so a point of $\Sigma\Sigma^{-1}\mathfrak{L}$, *i.e.* a point of $\mathfrak{L}$ not O, is contained in ΣD.

8. *Outline of the proof.*

We have to prove that a point of $\mathfrak{L}$ not O lies in each of the regions R defined in x, y coordinates by

$$|x^3+x^2y-2xy^2-y^3| \leqslant 1, \quad |x^3-xy^2-y^3| \leqslant 1.$$

We assume as a hypothesis that no point of $\mathfrak{L}$ except O is in R and deduce a contradiction. The proof, in fact, gives a finite region such that the the part of it which lies in R contains a point of $\mathfrak{L}$.

We apply Lemma 5 to several parallelograms, say $\mathfrak{P}_1$, $\mathfrak{P}_2$, ..., whose sides pass through some of the critical points $(0, \pm 1)$, $(\pm 1, 0)$, $(\pm 1, \pm 1)$, and whose areas lie mostly in R. These points $(0, \pm 1)$, etc., lead to the critical lattices for which the equality sign is required in the inequalities (10) and (11). We are then left with small parts of $\mathfrak{P}_1$, $\mathfrak{P}_2$, ..., say R_1, R_2, ..., which are not contained in R, and which must contain points of $\mathfrak{L}$, say, P_1, P_2, ... Various considerations enable us to reduce R_1, R_2, ... to smaller domains D_1, D_2, ..., *e.g.* the fact that $P_1 \pm P_2$, if not O, cannot be a point of R. The regions D_1, D_2, ... may lie near to and include one of the critical points $(0, \pm 1)$, etc. When, however, one of the critical points and its neighbourhood can be excluded from one of the D regions, a considerable advance has really been made in the solution of the problem. A prominent part is played by linear substitutions which transform the region R into itself and by the application of Lemma 6. We finally show that the sum or difference of two of the points P_1, P_2, ... lies in R, and this is contrary to the hypothesis*.

9. *Proof of Theorem* 4.

The region† R is defined in x, y coordinates by

$$|h(x, y)| = |x^3-xy^2-y^3| \leqslant 1. \tag{16}$$

---

* I am greatly indebted to Dr. Davenport for having verified my proof and for comments on the manuscript. I am also greatly indebted to Dr. Mahler for similar services and for drawing the figures.

† Fig. 1 will be found useful in reading the proof.

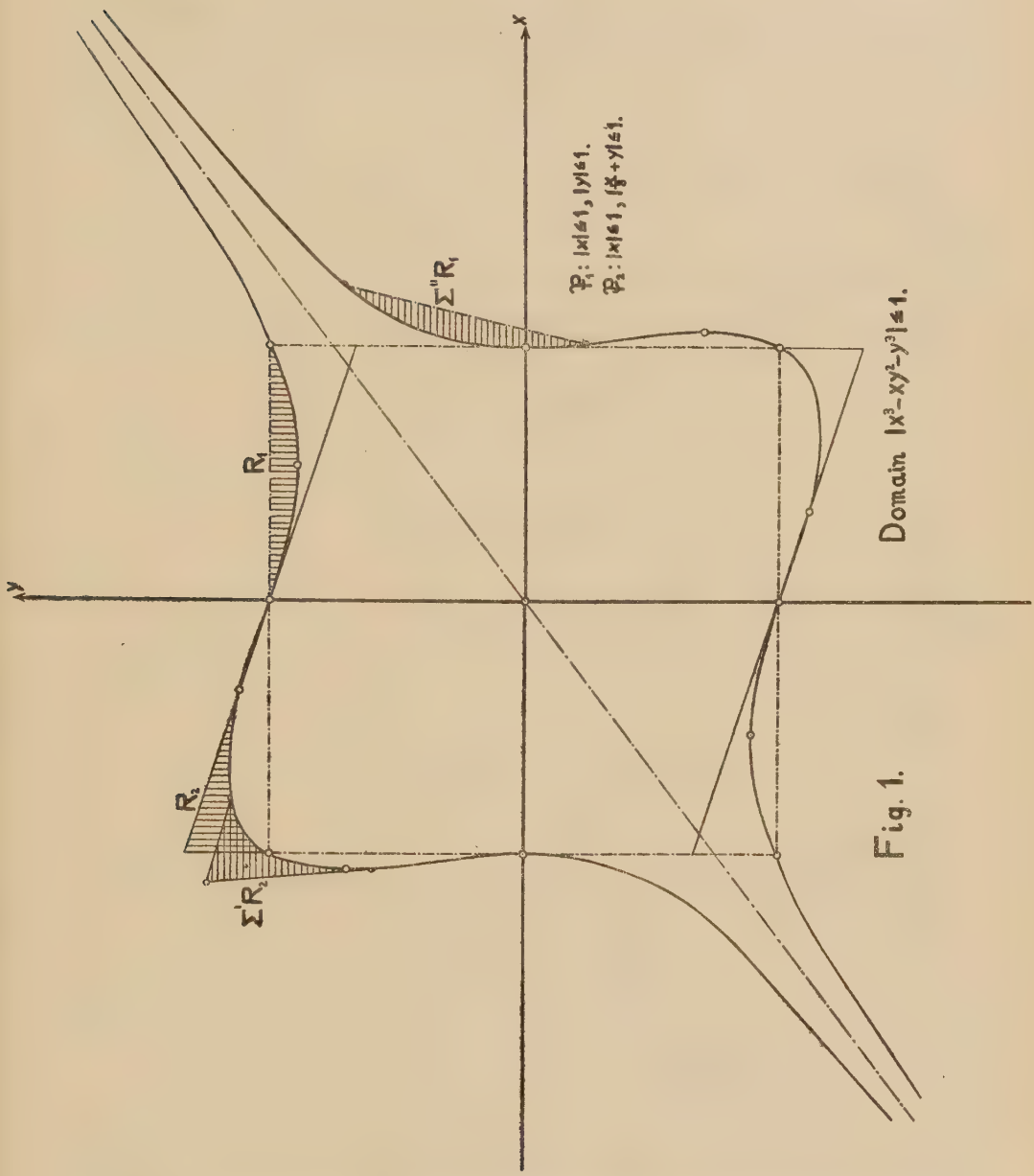


Fig. 1.

It is symmetrical about O and has for the upper and lower halves of its boundary

$$x^3-xy^2-y^3 = \mp 1,$$

respectively. The boundary passes through the vertices and mid-points

of the sides of the square $|x| = 1$, $|y| = 1$, and touches the sides $x = \pm 1$ at $(\pm 1, 0)$. The region R extends to infinity on both sides of the asymptote

$$x - \theta y = 0,$$

where $$\theta^3 - \theta - 1 = 0.$$

Also $$1{\cdot}324 < \theta < 1{\cdot}325, \tag{17}$$

which will, as usual, be written as $\theta = 1{\cdot}324\ldots$. More exactly, $\theta = 1{\cdot}3247\ldots$. Also

$$\theta^2 = 1{\cdot}7548\ldots, \quad 3\theta^2 - 1 = 4{\cdot}264\ldots. \tag{18}$$

The tangent at (0, 1) is given by

$$x + 3y - 3 = 0,$$

since $$(2xy + 3y^2)\frac{dy}{dx} - 3x^2 + y^2 = 0.$$

It meets the upper boundary again at $(-{\cdot}36, 1{\cdot}12)$. For, from

$$x^3 - x(1 - \tfrac{1}{3}x)^2 - (1 - \tfrac{1}{3}x)^3 = -1,$$

we have

$$25x^3 + 9x^2 = 0, \quad x = -\tfrac{9}{25} \quad \text{and} \quad y = \tfrac{28}{25}.$$

The tangent meets $x = -1$ where $y = \frac{4}{3}$.

The smallest ordinate in the first quadrant with $0 \leqslant x \leqslant 1$ is at the point $({\cdot}5\ldots, {\cdot}89\ldots)$. For $dy/dx = 0$ gives $y = x\sqrt{3}$,

$$\frac{y^3}{3\sqrt{3}} - \frac{y^3}{\sqrt{3}} - y^3 = -1,$$

$$y^3 = \frac{3\sqrt{3}}{3\sqrt{3}+2} = \frac{27 - 6\sqrt{3}}{23} = \frac{16{\cdot}607\ldots}{23} = {\cdot}722\ldots,$$

$$y = {\cdot}897\ldots \quad \text{and} \quad x = \frac{{\cdot}897\ldots}{1{\cdot}73\ldots} = {\cdot}5\ldots. \tag{19}$$

Hence y steadily decreases as x increases from 0 to ·5.

Some other boundary points, say (x, y), are important in the discussion. On the upper boundary

$$y^3 + xy^2 = x^3 + 1$$

these are $$x = \tfrac{1}{3},\ \cdot 17, \tag{20}$$

with ordinates $$y > \cdot 91,\ \cdot 94, \tag{21}$$

respectively. For

$$(\cdot 91)^3 + \tfrac{1}{3}(\cdot 91)^2 = \cdot 753571 + \tfrac{1}{3}(\cdot 8281) = \cdot 753571 + \cdot 2760\ldots = 1\cdot 029\ldots,$$

$$(\tfrac{1}{3})^3 + 1 = 1\cdot 037\ldots;$$

and

$$(\cdot 94)^3 + \cdot 17(\cdot 94)^2 = \cdot 830584 + \cdot 17(\cdot 8836) = \cdot 830584 + \cdot 150212 = \cdot 980796,$$

$$(\cdot 17)^3 + 1 = 1\cdot 004913.$$

It is also obvious that the ordinate $y > \cdot 91$ for $0 \leqslant x \leqslant \frac{1}{3}$.

The special points on the lower boundary

$$y^3 + xy^2 = x^3 - 1$$

are $$x = 1\cdot 25,\ 1\cdot 34, \tag{22}$$

with ordinates $$y < \cdot 70,\ \cdot 81, \tag{23}$$

respectively. For

$$(\cdot 7)^3 + 1\cdot 25(\cdot 7)^2 = \cdot 343 + \cdot 6125 = \cdot 9555,$$

$$(1\cdot 25)^3 - 1 = \cdot 953125;$$

also

$$(\cdot 81)^3 + 1\cdot 34(\cdot 81)^2 = \cdot 531441 + 1\cdot 34(\cdot 6561) = \cdot 531441 + \cdot 879174 = 1\cdot 410615.$$

$$(1\cdot 34)^3 - 1 = 1\cdot 406104.$$

It is also obvious that the ordinate $y < \cdot 81$ for $1 \leqslant x \leqslant 1\cdot 34$.

10. *Application of parallelograms of area* 4.

We now consider some parallelograms of area 4. The first is $\mathfrak{P}_1$, the square $|x| \leqslant 1$, $|y| \leqslant 1$, and it contains a point P_1, say, not O of $\mathfrak{L}$. All the points of $\mathfrak{P}_1$, are points of R except for a small region in the first quadrant defined by $y \leqslant 1$, $|h(x, y)| > 1$, and its image in the origin. Hence either P_1 is an inner point of R, or one of the eight boundary points of R, namely $(\pm 1, 0)$, $(0, \pm 1)$, $(\pm 1, \pm 1)$, $(\pm 1, \mp 1)$, or a point of the region R_1 defined by $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $|h(x, y)| \geqslant 1$, where $(0, 1)$, $(1, 1)$ may be

again included in R_1*. Every point of R_1 has an ordinate greater than ·89, and the ordinate of its curved boundary decreases steadily from 1 as x increases from 0 to ·5 ... and then increases steadily as x increases from ·5 ... to 1. It is clear from Lemma 4 that R_1 cannot contain two points of $\mathfrak{L}$ except when $\mathfrak{L}$ is $x = \xi$, $y = \eta$.

We may suppose that P_1 is not an inner point of R, since otherwise the main theorem is true. Suppose next that P_1 is one of the eight boundary points; by symmetry it suffices to consider (1, 0), (0, 1), (1, 1), (1, −1). We may suppose in this case that there is no other point of $\mathfrak{L}$ in R_1 or interior to R, as such a point would fall under the third or first alternative.

Take first (1, 0) to be a point of $\mathfrak{L}$. Then the lattice $\mathfrak{L}$ is defined by $x = \xi+\lambda\eta$, $y = \eta$, where $0 \leqslant \lambda < 1$, and ξ, η take all integer values. Then the point of $\mathfrak{L}$ given by taking $\eta = 1$, and ξ such that $0 \geqslant \xi+\lambda\eta > -1$ is an inner point of R unless $\lambda = 0$, $\xi = 0$; it then becomes a boundary point and $\mathfrak{L}$ is simply $x = \xi$, $y = \eta$. Then the equality sign is needed in (11).

Similarly, if (0, 1) is a point of $\mathfrak{L}$, the lattice is given by $x = \eta$, $y = \xi+\lambda\eta$, where $0 \leqslant \lambda < 1$ and ξ, η take all integer values. The point given by taking $\eta = 1$ and ξ such that $0 \leqslant \xi+\lambda\eta < 1$ is an inner point of R unless $\lambda = 0$, $\xi = 0$, and then $\mathfrak{L}$ is simply $x = \eta$, $y = \xi$.

In the remaining two cases when either (1, 1) or (1, −1) is a point of $\mathfrak{L}$, the points (1, 0), (0, 1) are by Lemma 5a, both points of $\mathfrak{L}$, *i.e.* $\mathfrak{L}$ is simply $x = \xi$, $y = \eta$.

Take as another parallelogram $\mathfrak{P}_2$ of area 4 that given by $|x| \leqslant 1$, $|y+\frac{1}{3}x| \leqslant 1$. Hence a lattice point of $\mathfrak{L}$, say P_2, either is an interior point of R or falls at $(0, \pm 1)$, $(\pm 1, 0)$, $(\pm 1, \mp 1)$, and all these need not be considered further, or P_2 is in the region R_2 consisting of the interior and sides of a triangle (with one curved side) whose vertices are at $(-\cdot 36, 1\cdot 12)$, $(-1, 1)$, $(-1, 1\frac{1}{3})$. The side joining the first two points is the corresponding part of the curve $x^3-xy^2-y^3 = -1$, and $(-1, 1)$ is again included in R_2.

R_2 cannot contain two lattice points, P_2, Q_2 say, for then the point P_2-Q_2 clearly is an inner point of R. For its coordinates x, y satisfy the inequalities $-1 < x < 1$, $-\frac{1}{3} \leqslant y \leqslant \frac{1}{3}$.

11. *Linear substitutions which change R into itself.*

LEMMA 7. *Two substitutions for which*

$$h(x', y') = h(x, y), \quad h(x'', y'') = -h(x, y),$$

* This is convenient later in applying Lemma 6.

are given by

$$\Sigma': \quad (3\theta^2-1)x' = -x-(\theta+3)y, \quad (3\theta^2-1)y' = -3\theta x+y, \tag{24}$$

and

$$\Sigma'': \quad x'' = -x', \quad y'' = -y', \tag{25}$$

where x', y' are given by (24).

For on factorising $h(x, y)$, we see that it suffices to take

$$x'-\theta y' = x-\theta y,$$

$$x'^2+\theta x'y'+\theta^{-1}y'^2 = x^2+\theta xy+\theta^{-1}y^2.$$

Hence, eliminating x', we have

$$(3\theta^2-1)y'^2+(\ldots)y' = 3\theta xy+(\theta^{-1}-\theta^2)y^2 = 3\theta xy-y^2.$$

Since one root is $y' = y$, the other is

$$(3\theta^2-1)y' = -3\theta x+y.$$

Then

$$(3\theta^2-1)x' = -3\theta^2 x+\theta y+(3\theta^2-1)(x-\theta y) = -x+(2\theta-3\theta^3)y.$$

The result follows since $\theta^3 = \theta+1$.

The substitutions are obviously self-inverse.

Since every lattice $\mathfrak{L}$, including $x = \xi$, $y = \eta$, having no point except O as an inner point of R, has a point in R_2, then by Lemma 6, the region $\Sigma' R_2$ contains a point $P_2'(x', y')$, not O, of $\mathfrak{L}$. If P_2 is $\Sigma'(-1, 1)$, the inscribing of the parallelogram $\Sigma'\mathfrak{P}$ in R shows as in § 10 that $\mathfrak{L}$ has a point other than O as an inner point of R unless $\mathfrak{L}$ is $\Sigma'(\xi, \eta)$. We can now exclude hereafter the cases $\mathfrak{L} = (\xi, \eta)$, $\mathfrak{L} = \Sigma'(\xi, \eta)$.

Lemma 8. $x_2 < -\cdot 77$.

The vertices of the region $\Sigma' R_2$ corresponding to the vertices

$$(-\cdot 36,\ 1\cdot 12), \quad (-1,\ 1), \quad (-1,\ \tfrac{4}{3}),$$

are

$$(-1\cdot 05\ldots,\ \cdot 59\ldots), \quad (-\cdot 77\ldots,\ 1\cdot 16\ldots), \quad (-1\cdot 11\ldots,\ 1\cdot 24\ldots). \tag{26}$$

For, with an obvious notation,

$$x'(-\cdot36,\ 1\cdot12) = \frac{\cdot36-1\cdot12(\theta+3)}{3\theta^2-1} = -\frac{4\cdot483\ldots}{4\cdot264\ldots} = -1\cdot05\ldots,$$

$$y'(-\cdot36,\ 1\cdot12) = \frac{1\cdot12+1\cdot08\theta}{3\theta^2-1} = \frac{2\cdot550\ldots}{4\cdot264\ldots} = \cdot59\ldots,$$

$$x'(-1,\ 1) = \frac{-\theta-2}{3\theta^2-1} = -\cdot77\ldots,$$

$$y'(-1,\ 1) = \frac{3\theta+1}{3\theta^2-1} = \frac{4\cdot974\ldots}{4\cdot264\ldots} = 1\cdot16\ldots,$$

$$x'(-1,\ \tfrac{4}{3}) = \frac{-3-\frac{4}{3}\theta}{3\theta^2-1} = -\frac{4\cdot766\ldots}{4\cdot264\ldots} = -1\cdot11\ldots,$$

$$y'(-1,\ \tfrac{4}{3}) = \frac{\frac{4}{3}+3\theta}{3\theta^2-1} = \frac{5\cdot307\ldots}{4\cdot264\ldots} = 1\cdot24\ldots.$$

Hence $$-1.12 < x' < -.77, \quad \cdot59 < y' < 1\cdot25.$$

Consider now the point $P_2 - P_2'$, say (x, y). Clearly

$$x < -\cdot36+1\cdot12 < \cdot8, \quad x > -1+\cdot77 = -\cdot23;$$

$$y < 1\cdot34-\cdot59 = \cdot75, \quad y > 1-1\cdot25 = -\cdot25.$$

Hence $P_2 - P_2'$ is an inner point of R, *i.e.* $P_2 = P_2'$, and P_2 is a common point of the regions R_2, $\Sigma' R_2$, and so $x_2 < -\cdot77$.

LEMMA 9. $x_1 < \cdot3$.

The area Δ of the triangle OP_1P_2 is either $\frac{1}{2}$ or 1. For

$$2\Delta = x_1y_2 - x_2y_1 \leqslant 1\tfrac{1}{3}+1 = 2\tfrac{1}{3},$$

and the result follows by Lemma 1.

We now prove that $\Delta = \frac{1}{2}$. For suppose that $\Delta = 1$. By Lemma 3, there must be a lattice point at the mid-point of P_1P_2. But then $\frac{1}{2}(P_1 - P_2)$ is a point (x, y) of $\mathfrak{L}$ which is an inner point of R since its coordinates satisfy

$$0 < x < 1, \quad 0 > y > \tfrac{1}{2}(\cdot89-\tfrac{4}{3}) > -\cdot23.$$

Hence

$$2\Delta = 1 = x_1y_2 - x_2y_1 > x_1 + (\cdot77 \times \cdot89) = x_1 + \cdot6853,$$

and so $x_1 < \frac{1}{3}$. But then $y_1 > \cdot91$ and so now

$$1 > x_1 + (\cdot77 \times \cdot91) = x_1 + \cdot7007,$$

whence $x_1 < \cdot3$.

12. *The transform of the region R_1.*

We consider the new region $\Sigma'' R_1$ deduced from R_1 by the substitution Σ''. The secant bounding the sector $\Sigma'' R_1$ has its vertices given by

$$\Sigma''(1, 1) = (1{\cdot}24\ldots, \ {\cdot}69\ldots), \quad \Sigma''(0, 1) = (1{\cdot}01\ldots, \ -{\cdot}23\ldots). \tag{27}$$

For $$x''(1, 1) = \frac{\theta+4}{3\theta^2-1}, \qquad y''(1, 1) = \frac{3\theta-1}{3\theta^2-1},$$

and so $$1{\cdot}24 < \frac{5{\cdot}324}{4{\cdot}265} < x''(1, 1) < \frac{5{\cdot}325}{4{\cdot}264} < 1{\cdot}25,$$

since $5{\cdot}325 < 5{\cdot}32625$;

$${\cdot}69 < \frac{2{\cdot}972}{4{\cdot}265} < y''(1, 1) < \frac{2{\cdot}975}{4{\cdot}264} < {\cdot}70,$$

since $2{\cdot}975 < 2{\cdot}9848$.

Also

$$x''(0, 1) = \frac{\theta+3}{3\theta^2-1} = 1{\cdot}01\ldots, \qquad y''(0, 1) = -\frac{1}{3\theta^2-1} = -{\cdot}23\ldots.$$

The secant above is derived from $y = 1$ by the transformation Σ'', and so its equation is

$$3\theta x - y = 3\theta^2 - 1.$$

It meets $y = 0$ where

$$x = \frac{3\theta^2-1}{3\theta} = 1{\cdot}07\ldots.$$

By Lemma 6, $\Sigma'' R_1$ contains a point $P_1''(x'', y'')$, say, of $\mathfrak{L}$ where clearly

$$1 < x'' < 1{\cdot}25, \quad -{\cdot}24 < y'' < {\cdot}70.$$

LEMMA 10. $y'' < 0$, $x'' < 1{\cdot}08$.

Consider $P(x, y)$, the point $P_2 + P_1'' - P_1$. Clearly

$$x < -{\cdot}77 + 1{\cdot}25 = {\cdot}48, \qquad x > -1 + 1 - {\cdot}3 = -{\cdot}3,$$

$$y > 1 - {\cdot}24 - 1 = -{\cdot}24, \qquad y < 1{\cdot}34 + {\cdot}70 - {\cdot}89 = 1{\cdot}15.$$

Hence there are three possibilities, of which the first is that P is a point of R other than O. This can be excluded and so we may suppose that $y > {\cdot}89$.

The second alternative is that P is not a point of R. Let $P - P_1$ be the point (X, Y). Then from the above,

$$X < {\cdot}48, \qquad X > -{\cdot}3 - {\cdot}3 = -{\cdot}6,$$

$$Y < 1{\cdot}15 - {\cdot}89 = {\cdot}26, \qquad Y > {\cdot}89 - 1 = -{\cdot}11.$$

Hence $P-P_1$ is an inner point of R and so $P = P_1$, whence

$$P_2+P_1'' = 2P_1.$$

We prove that this is impossible. It gives

$$y'' > 1{\cdot}78-1{\cdot}34 = {\cdot}44.$$

But since P_1 lies in the part of R_1 given by $x_1 < {\cdot}3$, by Lemma 6 and the transformation of $x_1 = {\cdot}3$ by Σ'', the point P'' in $\Sigma'' R_1$ also satisfies the inequality

$$x''+(\theta+3)y'' < {\cdot}3(3\theta^2-1),$$

i.e.

$$y'' < {\cdot}3\left(\frac{4{\cdot}265}{4{\cdot}324}\right) < {\cdot}3,$$

contradicting $y'' > {\cdot}44$.

Hence we are left with the third possibility, namely, that P is at the origin. Then

$$P_2+P_1'' = P_1,$$

$$y'' = y_1-y_2 < 0.$$

The other part $x'' < 1{\cdot}08$ is now obvious.

LEMMA 11. ${\cdot}17 < x_1 < {\cdot}26$.

Since $y'' < 0$, the transformation of $\Sigma'' R_1$ into R_1 shows that since only one point of $\mathfrak{L}$ lies in R_1, namely P_1, its coordinates, by Lemma 6, must satisfy the inequality arising from the transform of $y'' < 0$. Hence

$$(3\theta x_1-y_1)/(3\theta^2-1) < 0,$$

i.e.

$$x_1 \leqslant 1/(3\theta) < {\cdot}26.$$

Next $x_1 > {\cdot}17$. For if $x_1 \leqslant {\cdot}17$, the coordinates x, y of P_1+P'' satisfy

$$1 < x < {\cdot}17+1{\cdot}08 = 1{\cdot}25, \quad 1 > y > {\cdot}94-{\cdot}24 = {\cdot}70. \tag{28}$$

Hence P_1+P_1'' would be an inner point of R since the ordinate of the curve for $x = 1{\cdot}25$ is ${\cdot}69\ldots$.

LEMMA 12. $y'' > -{\cdot}08$.

Since $x_1 > {\cdot}17$, the transform of R_1 into $\Sigma'' R_1$ shows that y'' satisfies

$$y'' = \frac{3\theta x-y}{3\theta^2-1} > \frac{{\cdot}51\theta-1}{3\theta^2-1} = -\frac{{\cdot}324\ldots}{4{\cdot}264\ldots} > -{\cdot}08.$$

THEOREM 4. *The point P_1+P_1'' is an inner point of R.*

For its coordinates x, y satisfy the inequalities

$$1{\cdot}17 < x < {\cdot}26 + 1{\cdot}08 = 1{\cdot}34, \quad 1 > y > {\cdot}91 - {\cdot}08 = {\cdot}83.$$

But for $x = 1{\cdot}34$, the ordinate of the curve satisfies $y < {\cdot}81$.

This contradicts the assumption that R contains no point of $\mathfrak{L}$ except O. Hence Theorem 4 holds.

The proof shows that R must contain a point of $\mathfrak{L}$ lying in one of the following rectangles arising in discussing the left-hand side regions below:

(1) $\mathfrak{P}_1$	from which	$-1 \leqslant x \leqslant 1, \quad -1 \leqslant y \leqslant 1,$
§ 10 (2) R_2	from which	$-1 \leqslant x \leqslant 0, \quad 1 \leqslant y \leqslant 1{\cdot}34;$
§ 11 (3) $\Sigma' R_2$	from which	$-1{\cdot}12 \leqslant x \leqslant -{\cdot}77, \quad {\cdot}59 \leqslant y \leqslant 1{\cdot}25;$
§ 11 (4) $R_2 - \Sigma' R_2$	from which	$-{\cdot}23 \leqslant x \leqslant 1{\cdot}12, \quad -{\cdot}25 \leqslant y \leqslant {\cdot}75;$
§ 12 (5) $\Sigma'' R_1$	from which	$1{\cdot}01 < x < 1{\cdot}25, \quad -{\cdot}24 \leqslant y < {\cdot}70;$
§ 12 (6) $R_2 + \Sigma'' R_1 - R_1$	from which	$-{\cdot}3 < x < {\cdot}48, \quad -{\cdot}24 < y < 1{\cdot}15;$
§ 12 (7) $R_2 + \Sigma'' R_1 - 2R_1$	from which	$-1{\cdot}3 < x < 1{\cdot}48, \quad -1{\cdot}24 < y < {\cdot}30;$
§ 12 (8) $R_1 + \Sigma'' R_1$	from which	$1{\cdot}01 < x < 2{\cdot}25, \quad {\cdot}60 < y < 1{\cdot}70.$

By considering these results, we see that R must contain a point of $\mathfrak{L}$ lying in the rectangle

$$-1{\cdot}12 < x < 2{\cdot}25, \quad -1{\cdot}24 < y < 1{\cdot}70.$$

This point is easily found, and so the proof of the paper gives an actual construction of points of $\mathfrak{L}$ satisfying Theorem 4.

13. *Proof of Theorem* 3—*the region* R.

The region R is defined in x, y coordinates by

$$|g(x, y)| = |x^3 - x^2 y - 2xy^2 + y^3| \leqslant 1.$$

It is symmetrical about O. The boundary passes through the corners and the midpoints of the sides of the square $|x| \leqslant 1$, $|y| \leqslant 1$, and consists of six separate arcs, three from $g(x, y) = 1$ and three from $g(x, y) = -1$. Each arc is asymptotic to two of the three asymptotes

$$x + \theta y = 0, \quad x + \phi y = 0, \quad x + \psi y = 0,$$

and encloses a convex region since the real inflexions of the boundary are at infinity, as is obvious from the homogeneous form

$$XY(X + Y) - Z^3 = 0.$$

Here

$$\theta = 1{\cdot}2469\ldots, \quad \phi = -{\cdot}4450\ldots, \quad \psi = -1{\cdot}8019\ldots, \tag{29}$$

also

$$\theta^{-1} = {\cdot}8019\ldots, \quad \phi^{-1} = -2{\cdot}247\ldots, \quad \psi^{-1} = -{\cdot}5549\ldots. \tag{30}$$

The numbers θ, ϕ, ψ are well known in connection with the seventh roots of unity. Thus, if $t = e^{2\pi i/7}$,

$$\theta = t+t^{-1}, \quad \phi = t^2+t^{-2}, \quad \psi = t^3+t^{-3},$$

and

$$\phi\psi = \theta+\phi, \quad \psi\theta = \phi+\psi, \quad \theta\phi = \psi+\theta. \tag{31}$$

We require the intersections of the lines $x = 1$, $y = 1$, $y = x \pm 1$ with the boundary. These are given by the scheme

$x = 1$, $g(x, y) = -1$; $\quad y = -1, 1, 2.$

$x = 1$, $g(x, y) = 1$; $\quad y = 0, 1 \pm \sqrt{2}.$

$y = 1$, $g(x, y) = -1$; $\quad x = -1, 0, 2.$

$y = 1$, $g(x, y) = 1$; $\quad x = 1, \pm\sqrt{2}.$

$y = x+1$, $g(x, y) = -1$; $\quad (-2, -1), (-1, 0), (1, 2).$

$y = x+1$, $g(x, y) = 1$; $\quad (0, 1), (\sqrt{2}-1, \sqrt{2}), (-\sqrt{2}-1, -\sqrt{2}).$

$y = x-1$, $g(x, y) = -1$; $\quad (0, -1), (-\sqrt{2}+1, -\sqrt{2}), (\sqrt{2}+1, \sqrt{2}).$

$y = x-1$, $g(x, y) = 1$; $\quad (-1, -2), (1, 0), (2, 1).$

The tangents to the boundary are parallel to the x axis at the points

$$(-1{\cdot}17\ldots, -{\cdot}96\ldots), \quad ({\cdot}46\ldots, -{\cdot}84\ldots); \tag{32}$$

and perpendicular to the x axis at the points

$$({\cdot}96\ldots, -{\cdot}20\ldots), \quad ({\cdot}84\ldots, 1{\cdot}31\ldots); \tag{33}$$

and, of course, also at the images of these points in the origin. For, on differentiating, we have

$$(3y^2-4xy-x^2)\frac{dy}{dx}+3x^2-2xy-2y^2 = 0.$$

* See Fig. 2 in reading the proof.

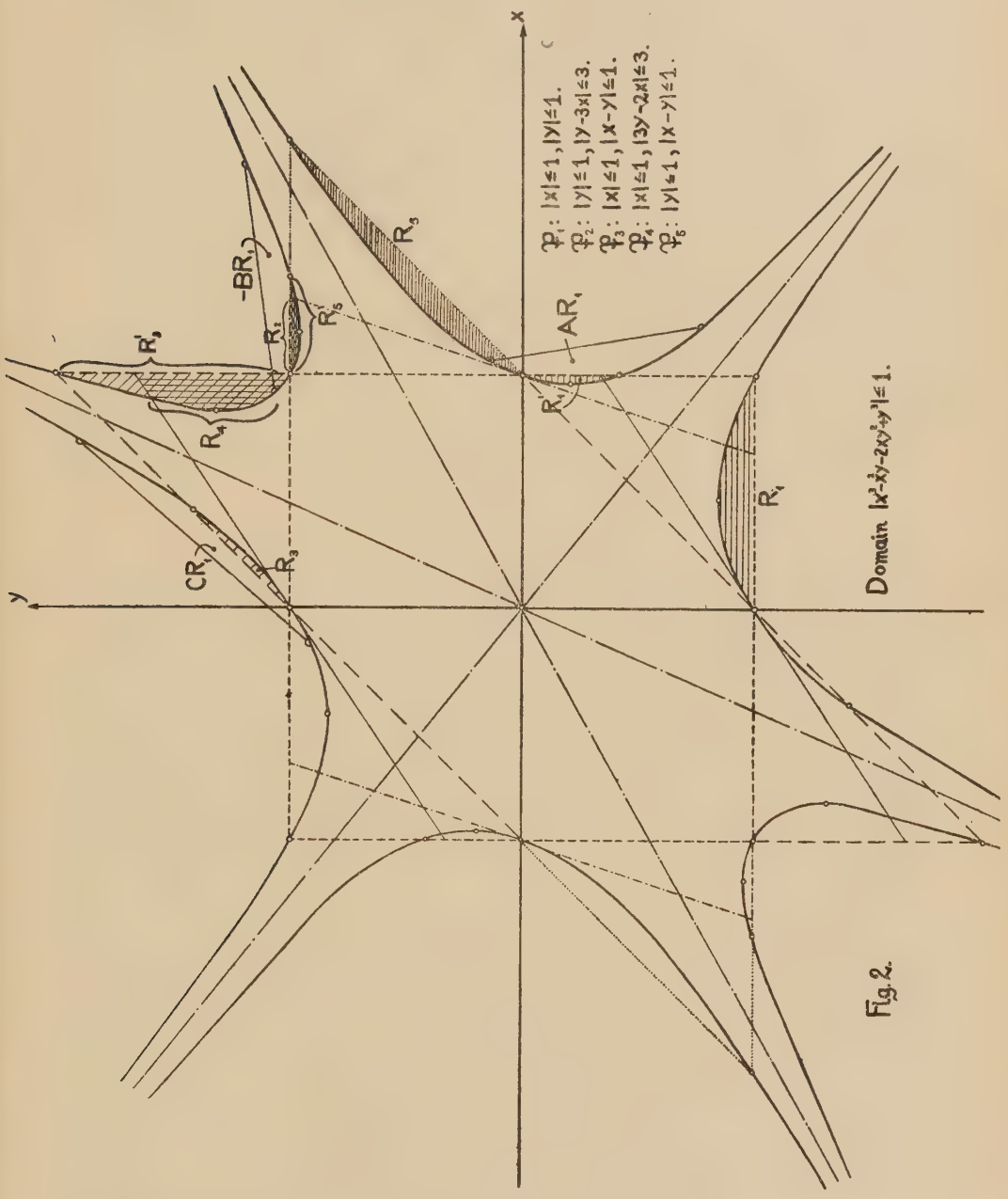


Fig. 2.

Thus $dy/dx = 0$ gives

$$3x^2 - 2xy - 2y^2 = 0, \quad 3x = y(1 \pm \sqrt{7}),$$

$$x = 1{\cdot}215\ldots y \quad \text{or} \quad -{\cdot}548\ldots y;$$

and $$-1{\cdot}1126\ldots y^3 = 1, \quad 1{\cdot}631\ldots y^3 = -1,$$

respectively, giving the above points.

Also $dx/dy = 0$ gives

$$3y^2 - 4yx - x^2 = 0, \quad x = y(-2 \pm \sqrt{7}),$$

$$x = -4{\cdot}645\ldots y \quad \text{or} \quad {\cdot}645\ldots y;$$

and $$-111{\cdot}5\ldots y^3 = 1, \quad -{\cdot}439\ldots y^3 = 1,$$

respectively, giving the above points.

We require the equations of the tangents at several points. These are given by the scheme

$$\left.\begin{array}{ll} (1,\,0)\colon\ y = 3(x-1), & (0,\,1)\colon\ y = \dfrac{2x}{3} + 1, \\ (1,\,1)\colon\ 2y + x - 3 = 0, & (1,\,-1)\colon\ 2y + x + 1 = 0. \end{array}\right\} \tag{34}$$

14. *Application of parallelograms of area* 4.

Consider the parallelogram $\mathfrak{P}_1$, $|x| \leqslant 1$, $|y| \leqslant 1$. This contains a point P_1, say, not O, of $\mathfrak{L}$. All the points of $\mathfrak{P}_1$ are points of R except for two small regions in the fourth quadrant and of course their images in the origin. Hence there are four possibilities: either P_1 is an inner point of R; or one of eight boundary points of R, namely $(\pm 1, 0)$, $(0, \pm 1)$, $(\pm 1, \pm 1)$, $(\pm 1, \mp 1)$; or a point of the region R_1' in the fourth quadrant defined by $|x| \leqslant 1$, $|y| \leqslant 1$, $|g(x, y)| \geqslant 1$ and lying near $(1, 0)$; or a point of the region R_1 in the fourth quadrant defined by $|x| \leqslant 1$, $|y| \leqslant 1$, $|g(x, y)| \geqslant 1$ and lying near the line $y = -1$ and where $(0, -1)$ and $(1, -1)$ are again included in R_1, and $(1, 0)$ in R_1'.

The first possibility has been excluded. Suppose next that P_1 is one of the eight boundary points. The procedure is similar to that in § 10. Thus if P_1 is $(\pm 1, 0)$, the lattice is $x = \xi + \lambda\eta$, $y = \eta$, where $0 \leqslant \lambda < 1$; and an inner point of R is given by taking $\eta = 1$, $0 \leqslant \xi + \lambda\eta < 1$, unless $\lambda = 0$, when it becomes a boundary point, and the lattice is $x = \xi$, $y = \eta$. Next, if P_1 is $(0, \pm 1)$, the lattice is $x = \eta$, $y = \xi + \lambda\eta$, $0 \leqslant \lambda < 1$, and an inner point of R is given by taking $\eta = 1$, $0 \leqslant \xi + \lambda\eta < 1$ unless $\lambda = 0$, etc. The argument for the boundary points $(1, \pm 1)$ is precisely the same as in § 10. We may suppose then that $\mathfrak{L}$ is not $x = \xi$, $y = \eta$.

The third and fourth possibilities mean that there may exist points of $\mathfrak{L}$, say P_1', P_1, in the respective regions R_1', R_1. Both cannot exist by Lemma 4 since they would lie in a square of area 1 with a vertex at O, and $\mathfrak{L}$ would be $x = \xi$, $y = \eta$, but one of them certainly exists.

Suppose first that the lattice is such that $P_1'(x_1', y_1')$ exists. Then since P_1' clearly lies below the tangent at (1, 0),

$$\cdot 96\ldots \leqslant x_1' \leqslant 1, \quad 1-\sqrt{2} \leqslant y_1' < 0, \quad y_1' < 3x_1' - 3. \tag{35}$$

Take now the parallelogram $\mathfrak{P}_2$ defined by $|y| \leqslant 1$, and $|y-3x| \leqslant 3$. Since no points of $\mathfrak{L}$ except O are contained in R or in R_1, there must be a point $P_2(x_2, y_2)$ of $\mathfrak{L}$ in the region R_2 of the first quadrant near (1, 1) defined by $y \leqslant 1$, $y-3x+3 \geqslant 0$, $|g(x, y)| \geqslant 1$. Clearly

$$1 < x_2 \leqslant \tfrac{4}{3}, \quad \cdot 96\ldots \leqslant y_2 \leqslant 1, \quad y_2 - 3x_2 + 3 \geqslant 0. \tag{36}$$

Take lastly the parallelogram $\mathfrak{P}_3$ defined by $|x| \leqslant 1$, $|x-y| \leqslant 1$. This defines two regions lying in the triangle, vertices (0, 1), (1, 1), (1, 2), one R_3 lying near (0, 1) and the other R_3' lying near $x = 1$; and these may contain lattice points P_3, P_3', say, of which, by Lemma 4, since $\mathfrak{L}$ is not $x = \xi$, $y = \eta$, only one can exist. This cannot be P_3'. For

$$2 \text{ area } OP_1' P_3' = x_1' y_3' - x_3' y_1' < 2+1.$$

Hence the area of the triangle $OP_1' P_3'$ is $\frac{1}{2}$ or 1. It cannot be 1, since then, by Lemma 3, the mid-point of $P_1' P_3'$ would be a point of $\mathfrak{L}$ which, since its ordinate satisfies $1 \geqslant y \geqslant 1-\frac{1}{2}\sqrt{2}$, would be a point of R. Hence the area is $\frac{1}{2}$ and

$$1 = x_1' y_3' - x_3' y_1' > \cdot 9 y_3'.$$

Then $y_3' < 1.12$ and so the point $P_2 - P_3'$ lies in the square $0 \leqslant x < 1$, $0 > y > -1$. Since its x coordinate is less than $\cdot 5$, it must be different from P_1', an impossibility by Lemma 4.

Hence P_3 exists and since P_3 lies above the tangent at (0, 1), and below the line $y = x+1$,

$$0 < x_3 \leqslant \sqrt{2}-1, \quad 1 < y_3 \leqslant \sqrt{2}, \quad y_3 > \tfrac{2}{3}x_3 + 1, \quad y_3 \leqslant x_3 + 1. \tag{37}$$

Hence we have three points P_1', P_2, P_3 of $\mathfrak{L}$, satisfying (35), (36), (37). We show now that this case is impossible.

LEMMA 13. *The point P_1' cannot exist.*

Consider the point $P_1' + P_3$, (x, y). We have

$$\cdot 96\ldots < x \leqslant \sqrt{2}, \quad 2-\sqrt{2} < y < \sqrt{2}.$$

Since obviously $y > x-1$, the point lies above the line $y = x-1$. By Lemma 4, it cannot lie between the curve and $y = x-1$ since there can be only one point of $\mathfrak{L}$, namely P_2, in the triangle with vertices (1, 0), (2, 1), (1, 1). Hence the point (x, y) will be a point of R if it lies below the

tangent at (1, 1), *i.e.* if

$$\tfrac{1}{2}x+y = y_1'+y_3+\tfrac{1}{2}(x_1'+x_3) < \tfrac{3}{2},$$

or, by (35) and (37), if

$$(3x_1'-3)+(x_3+1)+\tfrac{1}{2}(x_1'+x_3) < \tfrac{3}{2},$$

i.e.

$$x_1' < 1-\tfrac{3}{7}x_3.$$

Hence we may suppose

$$x_1' \geqslant 1-\tfrac{3}{7}x_3,$$

and then

$$x_1' \geqslant 1-\tfrac{3}{7}\cdot\tfrac{3}{2}(y_3-1) > 1-\tfrac{9}{14}\left(\frac{1}{x_1'}-1\right),$$

since $1 = x_1'y_3-x_3y_1' > x_1'y_3$. [For from (35), (37),

$$x_1'y_3-x_3y_1' < \sqrt{2}+(\sqrt{2}-1)^2 < 2.]$$

Hence

$$\tfrac{9}{14}\,\frac{1-x_1'}{x_1'} > (1-x_1'),$$

$$(\tfrac{9}{14}-x_1')(1-x_1') > 0,$$

which is false since $\cdot 9 \leqslant x_1' \leqslant 1$.

15. *A set of three points of the lattice.*

We have just proved that the lattice $\mathfrak{L}$ must be such that the only point of it other than O contained in $\mathfrak{P}_1$ is a point P_1 (and of course its image in the origin) in the region R_1 lying near $y = -1$. Clearly

$$0 \leqslant x_1 \leqslant 1, \quad -\cdot 84 \geqslant y_1 \geqslant -1. \tag{38}$$

We now find two more points of $\mathfrak{L}$. Take the parallelogram $\mathfrak{P}_4$ defined by $|x| \leqslant 1$, and $|y-\tfrac{2}{3}x| \leqslant 1$. By Lemma 5, there exists a point P_4 of $\mathfrak{L}$ in the region R_4 in the first quadrant defined by $y-\tfrac{2}{3}x-1 \leqslant 0$, $x \leqslant 1$, $|g(x, y)| > 1$. Clearly

$$x_4 > \cdot 84\ldots, \quad y_4 \leqslant 1\tfrac{2}{3}, \quad y_4 \leqslant \tfrac{2}{3}x_4+1. \tag{39}$$

Take, finally, $\mathfrak{P}_5$ defined by $|y| \leqslant 1$, $|y-x| \leqslant 1$. This defines two regions in the triangle vertices (1, 0), (2, 1), (1, 1), namely R_5' near $y = 1$, and R_5 near $y = x-1$, and these may contain lattice points P_5', P_5. By Lemma 4 exactly one of these must exist and this cannot be P_5'. For P_4-P_5' is an inner point of R since its coordinates x, y satisfy

$$\cdot 84\ldots-1\cdot 42 < x < 0, \quad 0 < y < 1\cdot 67-\cdot 96\ldots.$$

We now deduce sub-regions, say D_1, D_5, of R_1, R_5, respectively, containing the points P_1, P_5. For

$$2 \text{ area } OP_1P_4 = x_1y_4 - x_4y_1 < 1\tfrac{2}{3} + 1,$$

and so the area is $\frac{1}{2}$ or 1. We can exclude area 1, as then, by Lemma 3, the mid-point of P_1P_4 is a point of $\mathfrak{L}$ which is easily seen to be an inner point of R. Hence

$$1 = x_1y_4 - x_4y_1 > x_1 + \cdot 84 \times \cdot 84 > x_1 + \cdot 70.$$

Hence
$$x_1 < \cdot 30, \quad \text{and} \quad y_1 < -\cdot 84 \ldots \tag{40}$$

as already known.

Next
$$2 \text{ area } OP_1P_5 = x_1y_5 - x_5y_1 < \cdot 3 + 2,$$

and so the area is $\frac{1}{2}$ or 1. If the area is 1, then, by Lemma 3, the mid-point of P_1P_5 is a point of $\mathfrak{L}$. Then $\frac{1}{2}(P_5 - P_1)$ is also a point of $\mathfrak{L}$, but is obviously an inner point of R since its coordinates x, y satisfy

$$0 < x < 1, \quad \tfrac{1}{2} < y \leqslant 1.$$

Hence
$$1 = x_1y_5 - x_5y_1 > \cdot 84x_5,$$

and $x_5 < 1 \cdot 2$. Then since the ordinate of the boundary in R_5 increases as x increases from 1, $y_5 < \cdot 35$ from the equation $y^3 - 2xy^2 - x^2y + x^3 = 1$, which, for $x = 1 \cdot 2$, becomes $y^3 - 2 \cdot 4y^2 - 1 \cdot 44y + \cdot 728 = 0$. Hence

$$x_5 < 1 \cdot 2, \quad y_5 < \cdot 35. \tag{41}$$

16. *Linear transformations of R.*

We require linear transformations of the region R into itself. Its boundary is given by

$$|(x + \theta y)(x + \phi y)(x + \psi y)| = 1,$$

and so some are given by solving for x', y' in terms of x, y from the three equations, obviously equivalent to two,

$$(\phi - \psi)(x' + \theta y') = (\phi' - \psi')(x + \theta' y),$$
$$(\psi - \theta)(x' + \phi y') = (\psi' - \theta')(x + \phi' y),$$
$$(\theta - \phi)(x' + \psi y') = (\theta' - \phi')(x + \psi' y),$$

where θ', ϕ', ψ' is some permutation of θ, ϕ, ψ.

We pick out three, typified by $\theta' = \theta$, $\phi' = \psi$, $\psi' = \phi$, and refer to the corresponding transformations as A, B, C. Thus since, from (13),

$\psi(\phi-\theta)=\theta-\psi$, A is

$$x'+\theta y'=-x-\theta y,$$

$$x'+\phi y'=-\frac{x}{\psi}-y.$$

We can write the solution as

$$\left.\begin{aligned}(\theta-\phi)y'&=-\frac{x}{\psi}\,\frac{\theta}{\phi}-y\,\frac{\phi}{\psi}=-\theta^2x-\theta\phi^2y;\\ (\theta-\phi)x'&=-x\left(\frac{\theta}{\psi}-\phi\right)-y\theta(1-\phi)=x\,\frac{\phi}{\psi}+\psi y=\theta\phi^2x+\psi y.\end{aligned}\right\}\tag{42}$$

On noting that $\theta-\phi=\theta(\phi-\psi)$, we can write A in the form

$$\left.\begin{aligned}(\phi-\psi)x'&=\phi^2x+\phi\psi^2y,\\ (\phi-\psi)y'&=-\theta x-\phi^2y.\end{aligned}\right\}\tag{42'}$$

Similarly for B and C by cyclical interchange.

We now require the transforms by A, B, C of the small regions containing the point P_1. We begin with the sector R_1, whose transform by A is obviously the sector whose end points are the transforms by A of $(0, -1)$, $(1, -1)$, the end points of R_1.

Denote these transforms by $A\,(0, -1)$, etc., and call their coordinates $x[A(0, -1)]$, etc.

Then $A\,(0, -1)$ is

$$\left(\frac{-\psi}{\theta-\phi},\ \frac{\theta\phi^2}{\theta-\phi}\right),\tag{43}$$

which can be written as

$$\left(\frac{-\phi\psi^2}{\phi-\psi},\ \frac{\phi^2}{\phi-\psi}\right).$$

Also $A(1, -1)$ is

$$\left(\frac{\theta\phi^2-\psi}{\theta-\phi},\ \frac{-\theta^2+\theta\phi^2}{\theta-\phi}\right).\tag{43'}$$

An important result needed later is

$$A(0, -1)-C(0, -1)=A(1, -1).\tag{44}$$

For $$A(0, -1)-A(1, -1)=\left(-\frac{\theta\phi^2}{\theta-\phi},\ \frac{\theta^2}{\theta-\phi}\right),$$

and this is $C(0, -1)$ from the formula corresponding to the second form for $A(0, -1)$.

Lemma 14. *The equation of the transform of the line $y = -1$ by A is*

$$x(1-\psi)-y\phi = \theta-\psi.$$

For it is given in x', y' on eliminating x from

$$x'+\theta y' = -x+\theta, \quad x'+\phi y' = -\psi^{-1}x+1,$$

and is

$$x'(1-\psi)+y'(\theta-\phi\psi) = \theta-\psi,$$

whence the result follows from (31).

We now require approximate values for the coordinates of the points $\Sigma(0, -1)$, $\Sigma(1, -1)$, where $\Sigma = A, B, C$, as given by (42). Only crude approximations are really required.

We show that

$$\left.\begin{aligned} A(0, -1) &= (1{\cdot}06\ldots, {\cdot}14\ldots),\\ B(0, -1) &= (-{\cdot}91\ldots, -1{\cdot}06\ldots),\\ C(0, -1) &= (-{\cdot}14\ldots, {\cdot}91\ldots). \end{aligned}\right\} \tag{45}$$

From (31)

$$\theta\phi^2 = \phi\psi+\theta\phi = \theta+\phi+\psi+\theta = \theta-1;$$

hence

$$A(0, -1) = \left(-\frac{\psi}{\theta-\phi}, \frac{\theta-1}{\theta-\phi}\right) = \left(\frac{1{\cdot}801\ldots}{1{\cdot}692\ldots}, \frac{{\cdot}246\ldots}{1{\cdot}692\ldots}\right) = (1{\cdot}06\ldots, {\cdot}14\ldots);$$

$$B(0, -1) = \left(-\frac{\theta}{\phi-\psi}, \frac{\phi-1}{\phi-\psi}\right) = \left(-\frac{1{\cdot}246\ldots}{1{\cdot}356\ldots}, -\frac{1{\cdot}445\ldots}{1{\cdot}356\ldots}\right)$$

$$= (-{\cdot}91\ldots, -1{\cdot}06\ldots);$$

$$C(0, -1) = \left(-\frac{\phi}{\psi-\theta}, \frac{\psi-1}{\psi-\theta}\right) = \left(-\frac{{\cdot}445\ldots}{3{\cdot}048\ldots}, \frac{2{\cdot}801\ldots}{3{\cdot}048\ldots}\right)$$

$$= (-{\cdot}14\ldots, {\cdot}91\ldots).$$

We next show that

$$\left.\begin{aligned} A(1, -1) &= (1{\cdot}21\ldots, -{\cdot}77\ldots),\\ B(1, -1) &= (-1{\cdot}98\ldots, -1{\cdot}21\ldots,\\ C(1, -1) &= ({\cdot}77\ldots, 1{\cdot}98\ldots). \end{aligned}\right\} \tag{46}$$

For

$$A(1, -1) = \left(\frac{\theta\phi^2-\psi}{\theta-\phi}, \frac{-\theta^2+\theta\phi^2}{\theta-\phi}\right),$$

and from (31) and $\theta^2 = \phi+2$, this can be written as

$$A(1,-1) = \left(\frac{\theta-\psi-1}{\theta-\phi}, \frac{\theta-\phi-3}{\theta-\phi}\right) = \left(-1+\frac{2\theta-\phi-\psi-1}{\theta-\phi}, 1-\frac{3}{\theta-\phi}\right)$$

$$= \left(-1+\frac{3\theta}{\theta-\phi}, 1-\frac{3}{\theta-\phi}\right) = \left(-1+\frac{3\cdot740\ldots}{1\cdot692\ldots}, 1-\frac{3}{1\cdot692\ldots}\right)$$

$$= (1\cdot21\ldots, -\cdot77\ldots);$$

$$B(1,-1) = \left(-1+\frac{3\phi}{\phi-\psi}, 1-\frac{3}{\phi-\psi}\right) = \left(-1-\frac{1\cdot335\ldots}{1\cdot356\ldots}, 1-\frac{3}{1\cdot356\ldots}\right)$$

$$= (-1\cdot98\ldots, -1\cdot21\ldots);$$

$$C(1,-1) = \left(-1+\frac{3\psi}{\psi-\theta}, 1-\frac{3}{\psi-\theta}\right) = \left(-1+\frac{5\cdot405\ldots}{3\cdot048\ldots}, 1+\frac{3}{3\cdot048\ldots}\right)$$

$$= (\cdot77\ldots, 1\cdot98\ldots).$$

We now require the transform of the part of R_1 given by $x < \cdot3$. The end points of the vertical boundary are $(\cdot3, -1)$ and $(\cdot3, -\cdot86\ldots)$. One significant figure will suffice. We show that

$$\left.\begin{aligned} A(\cdot3,-1) &= (1\cdot1\ldots, -\cdot1\ldots);\\ B(\cdot3,-1) &= (-1\cdot2\ldots, -1\cdot1\ldots);\\ C(\cdot3,-1) &= (\cdot1\ldots, 1\cdot2\ldots). \end{aligned}\right\} \qquad (47)$$

In vector notation

$$(\cdot3,-1) = \cdot7(0,-1)+\cdot3(1,-1),$$

and so from (45), (46),

$$A(\cdot3,-1) = \cdot7(1\cdot06\ldots, \cdot14\ldots)+\cdot3(1\cdot21\ldots, -\cdot77\ldots) = (1\cdot1\ldots, -\cdot1\ldots);$$

$$B(\cdot3,-1) = \cdot7(-\cdot91\ldots, -1\cdot06\ldots)+\cdot3(-1\cdot98\ldots, -1\cdot21\ldots)$$

$$= (-1\cdot2\ldots, -1\cdot1\ldots);$$

$$C(\cdot3,-1) = \cdot7(-\cdot14\ldots, \cdot91\ldots)+\cdot3(\cdot77\ldots, 1\cdot98\ldots) = (\cdot1\ldots, 1\cdot2\ldots).$$

We next show

$$\left.\begin{aligned} A(\cdot3,-\cdot86\ldots) &= (\cdot9\ldots, -\cdot1\ldots),\\ B(\cdot3,-\cdot86\ldots) &= (-1\cdot0\ldots, -\cdot8\ldots),\\ C(\cdot3,-\cdot86\ldots) &= (\cdot1\ldots, 1\cdot0\ldots). \end{aligned}\right\} \qquad (48)$$

In vector notation

$$(\cdot3,-\cdot86\ldots) = (\cdot56\ldots)(0,-1)+\cdot3(1,-1),$$

and so from (45), (46),

$$A(\cdot 3, -\cdot 86\ldots) = (\cdot 56\ldots)(1\cdot 06\ldots, \cdot 14\ldots)+\cdot 3(1\cdot 21\ldots, -\cdot 77\ldots)$$
$$= (\cdot 9\ldots, -\cdot 1\ldots);$$
$$B(\cdot 3, -\cdot 86\ldots) = (\cdot 56\ldots)(-\cdot 91\ldots, -1\cdot 06\ldots)+\cdot 3(-1\cdot 98\ldots, -1\cdot 21\ldots)$$
$$= (-1\cdot 0\ldots, -\cdot 8\ldots);$$
$$C(\cdot 3, -\cdot 86\ldots) = (\cdot 56\ldots)(-\cdot 14\ldots, \cdot 91\ldots)+\cdot 3(\cdot 77\ldots, 1\cdot 98\ldots)$$
$$= (\cdot 1\ldots, 1\cdot 0\ldots).$$

17. *Bounds for the three points P_1, P_4, P_5.*

It has now been shown that every lattice $\mathfrak{L}$, of which no point except O is an inner point of R, must have three of its points P_1, P_4, P_5 in the respective regions R_1, R_4 R_5, obvious end points arising from $x=\xi$, $y=\eta$. By Lemma 6, there must be points of $\mathfrak{L}$, say $A'P_1$, $B'P_1$, $C'P_1$, in the respective regions AR_1, BR_1, CR_1. The notation $A'P_1$ does not mean the transform of P_1 by A. The points $A(0, -1)$, $A(1, -1)$, etc. are easily dealt with. For as in § 14, if any of these is a point of $\mathfrak{L}$, the parallelogram $A\mathfrak{P}_1$, etc. shows that a point of $\mathfrak{L}$ not O is an inner point of R except when $\mathfrak{L}$ is the lattice $(A\xi, A\eta)$. We may suppose now that $\mathfrak{L}$ is not (ξ, η) or $(\Sigma\xi, \Sigma\eta)$.

Lemma 15. $A'P_1 = P_5$.

Call $A'P_1$ the point (x, y). The sector AR_1 extends from $A(0, -1)$ to $A(1, -1)$, say $(1\cdot 06\ldots, \cdot 14\ldots)$ to $(1\cdot 21\ldots, -\cdot 77\ldots)$. Hence $x>1$, since AR_1 contains the region R_1', really the part of AR_1 with $x\leqslant 1$, which has been shown to contain no point of $\mathfrak{L}$. Now $A'P_1$ must also lie in the A transform of D_1, *i.e.*, by (47), (48), in a curvilinear triangle whose vertices are at

$$(1\cdot 06\ldots, \cdot 14\ldots), \quad (\cdot 9\ldots, -\cdot 1\ldots), \quad (1\cdot 1\ldots, -\cdot 1\ldots),$$

the sides through $(1\cdot 1\ldots, -\cdot 1\ldots)$ being straight lines.

But from (41),

$$1<x_5<1\cdot 2, \quad 0<y_5<\cdot 35,$$

and then $P_5-A'P_1$ is clearly an inner point of R and must be O.

It proves convenient to divide into two parts the region AD_1 containing P_5.

The first is given by

$$x_5 \leqslant x[A(0, -1)] = 1{\cdot}06\ldots, \quad y_5 \leqslant y[A(0, -1)] = {\cdot}14\ldots. \tag{49}$$

The second lies beneath the transform by A of the chord $y = -1$ bounding the sector R_1 whose equation, by Lemma 14, is

$$x(1-\psi) - y\phi = \theta - \psi.$$

It meets the x axis where

$$x = \frac{\theta-\psi}{1-\psi} = \frac{3{\cdot}048\ldots}{2{\cdot}801\ldots} = 1{\cdot}08\ldots.$$

Hence x_5, y_5 satisfy the inequality

$$y_5 < \frac{\theta-\psi}{-\phi} + x_5\,\frac{1-\psi}{\phi}.$$

This inequality is also satisfied when (x_5, y_5) lies in the first part of AD_1. We use it only in the crude form

$$y_5 < 7 - 6x_5.$$

LEMMA 16. $B' P_1 = -P_4$.

By Lemma 6, a point of $\mathcal{L}$ lies in the sector $-BR_1$, *i.e.* the image of BR_1 in O, extending from $({\cdot}91\ldots, 1{\cdot}06\ldots)$ to $(1{\cdot}98\ldots, 1{\cdot}21\ldots)$. There are three possibilities for this point $-B' P_1$, say (x, y).

First $y \leqslant 1$, a hypothesis which can be rejected. For then $x \geqslant 1$, but $-B' P_1$ cannot be P_5 since $y_5 < {\cdot}35$, and by Lemma 4, $-B' P_1$ cannot coexist with P_5.

Secondly, it lies in R_4 and must then coincide with P_4. Then $-B' P_1 = P_4$, and also, from the coordinates of $-BR_1$,

$$x_4 > {\cdot}91\ldots, \quad y_4 < 1{\cdot}21\ldots. \tag{50}$$

Thirdly, $x \geqslant 1$, $y \geqslant 1$. This leads to a contradiction and can be rejected. For from (39), $x_4 > {\cdot}8$, $y_4 \leqslant 1\frac{2}{3}$, and so $x > 1{\cdot}8$ since if $x \leqslant 1{\cdot}8$, $-B' P_1 - P_4$ is an inner point of R. But $x < 1{\cdot}2\ldots$ from (47), (48) since $-B' P_1$ must also lie in the $-B$ transform of D_1.

LEMMA 17. $C' P_1 = -P_1$.

By Lemma 6, a point of $\mathcal{L}$, $C' P_1$, say (x, y), lies in CR_1, a sector extending from $(-{\cdot}14\ldots, {\cdot}91\ldots)$ to $({\cdot}77\ldots, 1{\cdot}98\ldots)$, and so lies in either the first or second quadrants.

In the first case, it also lies in CD_1, and so from (47), (48),

$$0 \leqslant x < \cdot 1 \ldots, \quad 1 \leqslant y < 1 \cdot 2 \ldots.$$

Then from (50), $P_4 - C' P_1$ is certainly not 0. It either lies in the first quadrant and is then an inner point of R; or lies in the fourth quadrant but is not P_1 (since its abscissa is greater than $\cdot 7$), an impossibility by Lemma 4.

Hence $C' P_1$ must lie in the second quadrant and must coincide with $-P_1$ from Lemma 4. Hence $C' P_1 = -P_1$ and then also

$$\left.\begin{aligned} x_1 &< -x[C(0, -1)] = x[C(0, 1)] < \cdot 14 \ldots, \\ y_1 &< -y[C(0, -1)] = y[C(0, 1)] < -\cdot 91 \ldots. \end{aligned}\right\} \tag{51}$$

18. *Final stage in proof of Theorem* 3.

Theorem 3. *$P = P_1 + P_5$ is a point of R.*

Its coordinates x, y are given by

$$x = x_1 + x_5, \quad y = y_1 + y_5.$$

Then from (51) and § 17,

$$\left.\begin{aligned} x &< x[C(0, 1)] + x_5 = \cdot 14 + x_5, \\ y &< y[C(0, 1)] + y_5 < y[C(0, 1)] + \frac{\theta - \psi}{-\phi} + x_5 \frac{1 - \psi}{\phi} \\ &< -\cdot 9 + 7 - 6x_5 = 6 \cdot 1 - 6x_5. \end{aligned}\right\} \tag{52}$$

Obviously $x > 1$, $y > -1$.

We have to show that P is a point of R when

$$1 < x_5 < \frac{\theta - \psi}{1 - \psi} = 1 \cdot 08 \ldots.$$

Suppose first that

$$x_5 < x[A(0, -1)] = 1 \cdot 06 \ldots, \quad y_5 < y[A(0, -1)] = \cdot 14 \ldots.$$

Clearly from (44),

$$\left.\begin{aligned} x &< x[C(0, 1)] + x[A(0, -1)] = x[A(1, -1)] = 1 \cdot 21 \ldots, \\ y &< y[C(0, 1)] + y[A(0, -1)] = y[A(1, -1)] = -\cdot 77 \ldots. \end{aligned}\right\} \tag{53}$$

But $y[A(1, -1)]$ is the ordinate at $x[A(1, -1)]$, and since $y > -1$, P lies in R.

Suppose next that

$$1 \cdot 06 \ldots = x[A(0, -1)] < x_5 < \frac{\theta - \psi}{1 - \psi} = 1 \cdot 08 \ldots. \tag{54}$$

We prove that P lies below the tangent at $A(1, -1)$, and since $y > -1$, P is again an inner point of R. We now require the equation of the tangent at $A(1, -1)$. It suffices for the argument below to find the approximate equation of the tangent at $(1{\cdot}21\ldots, -{\cdot}77\ldots)$. This is

$$(y+{\cdot}77)+\frac{5{\cdot}05}{4{\cdot}06}(x-1{\cdot}21)=0, \tag{55}$$

since

$$\frac{dy}{dx}=\frac{3x^2-2xy-2y^2}{x^2+4xy-3y^2}=\frac{4{\cdot}38+1{\cdot}86-1{\cdot}19}{1{\cdot}46-3{\cdot}72-1{\cdot}78}\ \text{(approximately)}=-\frac{5{\cdot}05}{4{\cdot}04}.$$

The condition for the point x, y to be below the line (55) is that the left-hand side of (55) is negative. Substituting from (52), we have roughly

$$5{\cdot}5-5x_5<0.$$

Hence on substituting for x, y in terms of x_5 from (52), the condition for x, y to be below the tangent at $A(1, -1)$ takes the shape*

$$t\big(x[A(0, -1)]-x_5\big)<0, \tag{56}$$

where t is a positive number nearly 5, and the $x[A(0, -1)]$ arises since from (53) there must be equality when $x_5=x[A(0, -1)]$. The condition (56) is satisfied since $x_5>x[A(0, -1)]$.

This proves Theorem 3.

The point of $\mathfrak{L}$ lying in R thus found, also lies in the rectangle $|x|\leqslant 3{\cdot}4$, $|y|\leqslant 3{\cdot}7$. This is easily seen on taking crude bounds for the positions of the points P_1', P_2, P_3, $P_1'+P_3$, P_1, P_4, P_5, P_4+P_5, $P_5-A'P_1$, $P_4+B'P_1$, $P_4-C'P_1$.

[*Note added* 23 *March*, 1943, in reading the proofs. Shorter proofs of Theorems 3, 4 have since been found both by Davenport and myself and will appear in the *Journal* of the Society†.]

The University,
Manchester.

---

* Really $$y_5-y[A(0, -1)]-\frac{dy_5}{dx_5}\big(x_5-x[A(0, -1)]\big)<0$$

with $$y_5-y[A(0, -1)]<\frac{1-\psi}{\phi}\big(x_5-x[A(0, -1)]\big).$$

† I should like to thank Prof. Davenport and Dr. R. Rado for help with the proof sheets.

A TABLE OF PARTITIONS

By J. A. Todd.

[Received 10 December, 1940.—Read 21 May, 1942.]

1. Let $p(n)$ be the number of unrestricted partitions of the positive integer n, and let $p(n, m)$ denote the number of partitions of n into exactly m parts. Then

$$p(n, 1) = 1 \tag{1}$$

and, for $m > 1$,

$$p(n, m) = p(n-1, m-1)+p(n-m, m), \tag{2}$$

since the first term on the right of (2) represents the number of partitions of n into m parts, one at least of which is unity, and the second term represents the number of partitions of n into m parts, each of which exceeds unity. By using this formula recursively the values of $p(n, m)$ may be calculated for each pair of integers n, m. This leads to Euler's table of double entry, described by Chrystal†.

It is clear that $p(n, m) = 0$ if $m > n$, and that, if $m \geqslant \frac{1}{2}n$,

$$p(n, m) = p(n-m). \tag{3}$$

Thus it is sufficient to tabulate the values of $p(n, m)$ for $m \leqslant \frac{1}{2}n$.

The present table contains the values of $p(n, m)$ for $1 \leqslant n \leqslant 100$, $1 \leqslant m \leqslant [\frac{1}{2}(n+1)]$, the final entry, for each value of n, being $p[\frac{1}{2}(n+1)]$ corresponding to the value $m = [\frac{1}{2}(n+1)]$. This final entry is marked

† G. Chrystal, *Algebra*, 2 (1900), 557–561. Euler's table [*Opera* (1), 8 (1922), 327–328] gives $p(n+m, m)$ for $1 \leqslant m \leqslant 11$, $1 \leqslant n \leqslant 69$,

with an asterisk to indicate that it and the unentered values of $p(n, m)$ for larger values of m are obtainable from a table of unrestricted partitions by means of formula (3). For this purpose an extract from Gupta's table up to $n = 600$ is given on p. 233. The values of $p(n, m)$ were obtained from the formula

$$p(n, m) = \sum_{r=1}^{m} p(n-m, r), \tag{4}$$

which is an immediate consequence of (2). Since

$$\sum_{m=1}^{n} p(n, m) = p(n), \tag{5}$$

a check on the work is provided by the formula

$$\sum_{m=1}^{s} p(n, m) = p(n) - \sum_{r=0}^{n-s-1} p(r), \tag{6}$$

where $s = [\frac{1}{2}(n+1)]$, the expression on the right being computed from the values of $p(n)$ tabulated by Gupta†.

2. It is easily seen that the function

$$f_m(x) = \frac{x^m}{(1-x)(1-x^2)\dots(1-x^m)} \tag{7}$$

is the generating function of $p(n, m)$. It is easily seen from (2) that for fixed m, $p(n, m)$ behaves on the average like a polynomial of degree $m-1$ in n, in which the dominant term is

$$\frac{n^{m-1}}{(m-1)!\, m!}, \tag{8}$$

an expression which gives the approximate order of magnitude of $p(n, m)$ when n is large compared with m. Thus (8) gives the order of magnitude of $p(n, m)$ for fixed m, when n is large.

Trinity College,
Cambridge.

---

† H. Gupta, *Proc. London Math. Soc.* (2), 39 (1935), 142–149. We define, conventionally, $p(0) = 1$.

m \ n	1	2	3	4	5	6	7	8	9	10
1	*1	*1	1	1	1	1	1	1	1	1
2			*1	*2	2	3	3	4	4	5
3					*2	*3	4	5	7	8
4							*3	*5	6	9
5									*5	*7

m \ n	11	12	13	14	15	16	17	18	19	20
1	1	1	1	1	1	1	1	1	1	1
2	5	6	6	7	7	8	8	9	9	10
3	10	12	14	16	19	21	24	27	30	33
4	11	15	18	23	27	34	39	47	54	64
5	10	13	18	23	30	37	47	57	70	84
6	*7	*11	14	20	26	35	44	58	71	90
7			*11	*15	21	28	38	49	65	82
8					*15	*22	29	40	52	70
9							*22	*30	41	54
10									*30	*42

m \ n	21	22	23	24	25	26	27	28	29	30
1	1	1	1	1	1	1	1	1	1	1
2	10	11	11	12	12	13	13	14	14	15
3	37	40	44	48	52	56	61	65	70	75
4	72	84	94	108	120	136	150	169	185	206
5	101	119	141	164	192	221	255	291	333	377
6	110	136	163	199	235	282	331	391	454	532
7	105	131	164	201	248	300	364	436	522	618
8	89	116	146	186	230	288	352	434	525	638
9	73	94	123	157	201	252	318	393	488	598
10	55	75	97	128	164	212	267	340	423	530
11	*42	*56	76	99	131	169	219	278	355	445
12			*56	*77	100	133	172	224	285	366
13					*77	*101	134	174	227	290
14							*101	*135	175	229
15									*135	*176

m \ n	31	32	33	34	35	36	37	38
1	1	1	1	1	1	1	1	1
2	15	16	16	17	17	18	18	19
3	80	85	91	96	102	108	114	120
4	225	249	270	297	321	351	378	411
5	427	480	540	603	674	748	831	918
6	612	709	811	931	1057	1206	1360	1540
7	733	860	1009	1175	1367	1579	1824	2093
8	764	919	1090	1297	1527	1801	2104	2462
9	732	887	1076	1291	1549	1845	2194	2592
10	653	807	984	1204	1455	1761	2112	2534
11	560	695	863	1060	1303	1586	1930	2331
12	460	582	725	905	1116	1380	1686	2063
13	373	471	597	747	935	1158	1436	1763
14	293	378	478	608	762	957	1188	1478
15	230	295	381	483	615	773	972	1210
16	*176	*231	296	383	486	620	780	983
17			*231	*297	384	488	623	785
18					*297	*385	489	625
19							*385	*490

m \ n	39	40	41	42	43	44	45	46
1	1	1	1	1	1	1	1	1
2	19	20	20	21	21	22	22	23
3	127	133	140	147	154	161	169	176
4	441	478	511	551	588	632	672	720
5	1014	1115	1226	1342	1469	1602	1747	1898
6	1729	1945	2172	2432	2702	3009	3331	3692
7	2400	2738	3120	3539	4011	4526	5102	5731
8	2857	3319	3828	4417	5066	5812	6630	7564
9	3060	3589	4206	4904	5708	6615	7657	8824
10	3015	3590	4242	5013	5888	6912	8070	9418
11	2812	3370	4035	4802	5708	6751	7972	9373
12	2503	3036	3655	4401	5262	6290	7476	8877
13	2164	2637	3210	3882	4691	5635	6761	8073
14	1819	2241	2738	3345	4057	4920	5928	7139
15	1508	1861	2297	2815	3446	4192	5096	6158
16	1225	1530	1891	2339	2871	3523	4293	5231
17	990	1236	1545	1913	2369	2913	3579	4370
18	788	995	1243	1556	1928	2391	2943	3621
19	626	790	998	1248	1563	1939	2406	2965
20	*490	*627	791	1000	1251	1568	1946	2417
21			*627	*792	1001	1253	1571	1951
22					*792	*1002	1254	1573
23							*1002	*1255

m \ n	47	48	49	50	51	52
1	1	1	1	1	1	1
2	23	24	24	25	25	26
3	184	192	200	208	217	225
4	764	816	864	920	972	1033
5	2062	2233	2418	2611	2818	3034
6	4070	4494	4935	5427	5942	6510
7	6430	7190	8033	8946	9953	11044
8	8588	9749	11018	12450	14012	15765
9	10156	11648	13338	15224	17354	19720
10	10936	12690	14663	16928	19466	22367
11	11004	12866	15021	17475	20298	23501
12	10489	12384	14552	17084	19978	23334
13	9624	11424	13542	15988	18847	22142
14	8551	10232	12186	14499	17176	20325
15	7434	8932	10715	12801	15272	18148
16	6334	7665	9228	11098	13287	15892
17	5332	6469	7841	9459	11395	13671
18	4426	5409	6570	7976	9635	11626
19	3651	4468	5465	6647	8077	9770
20	2980	3673	4498	5507	6703	8154
21	2424	2991	3688	4520	5537	6745
22	1954	2429	2998	3699	4535	5559
23	1574	1956	2432	3003	3706	4546
24	*1255	*1575	1957	2434	3006	3711
25			*1575	*1958	2435	3008
26					*1958	*2436

n	$p(n)$	n	$p(n)$	n	$p(n)$	n	$p(n)$	n	$p(n)$
1	1	11	56	21	792	31	6842	41	44583
2	2	12	77	22	1002	32	8349	42	53174
3	3	13	101	23	1255	33	10143	43	63261
4	5	14	135	24	1575	34	12310	44	75175
5	7	15	176	25	1958	35	14883	45	89134
6	11	16	231	26	2436	36	17977	46	1 05558
7	15	17	297	27	3010	37	21637	47	1 24754
8	22	18	385	28	3718	38	26015	48	1 47273
9	30	19	490	29	4565	39	31185	49	1 73525
10	42	20	627	30	5604	40	37338	50	2 04226

m \ n	53	54	55	56	57	58
1	1	1	1	1	1	1
2	26	27	27	28	28	29
3	234	243	252	261	271	280
4	1089	1154	1215	1285	1350	1425
5	3266	3507	3765	4033	4319	4616
6	7104	7760	8442	9192	9975	10829
7	12241	13534	14950	16475	18138	19928
8	17674	19805	22122	24699	27493	30588
9	22380	25331	28629	32278	36347	40831
10	25608	29292	33401	38047	43214	49037
11	27169	31316	36043	41373	47420	54218
12	27156	31570	36578	42333	48849	56297
13	25971	30366	35452	41269	47968	55610
14	23961	28212	33104	38797	45326	52888
15	21535	25469	30073	35401	41612	48772
16	18928	22518	26694	31603	37292	43951
17	16380	19551	23303	27684	32839	38837
18	13968	16765	20040	23928	28472	33834
19	11802	14199	17062	20425	24418	29098
20	9871	11937	14375	17293	20722	24803
21	8210	9948	12038	14510	17469	20953
22	6775	8252	10004	12115	14611	17604
23	5574	6797	8282	10046	12171	14688
24	4553	5585	6812	8304	10076	12213
25	3714	4558	5592	6823	8319	10098
26	3009	3716	4561	5597	6830	8330
27	*2436	*3010	3717	4563	5600	6835
28			*3010	*3718	4564	5602
29					*3718	*4565

m \ n	59	60	61	62	63	64
1	1	1	1	1	1	1
2	29	30	30	31	31	32
3	290	300	310	320	331	341
4	1495	1575	1650	1735	1815	1906
5	4932	5260	5608	5969	6351	6747
6	11720	12692	13702	14800	15944	17180
7	21873	23961	26226	28652	31275	34082
8	33940	37638	41635	46031	50774	55974
9	45812	51294	57358	64015	71362	79403
10	55494	62740	70760	79725	89623	1 00654
11	61903	70515	80215	91058	1 03226	1 16792
12	64707	74287	85067	97299	1 11036	1 26560
13	64370	74331	85711	98609	1 13287	1 29883
14	61538	71509	82882	95943	1 10795	1 27786
15	57080	66634	77667	90316	1 04875	1 21510
16	51643	60603	70927	82898	96650	1 12540
17	45864	54012	63516	74506	87268	1 01982
18	40080	47420	55940	65907	77449	90889
19	34624	41078	48668	57503	67846	79855
20	29588	35251	41869	49668	58754	69414
21	25100	29973	35741	42496	50460	59755
22	21129	25331	30270	36126	42986	51087
23	17705	21264	25507	30501	36423	43371
24	14744	17782	21365	25642	30677	36654
25	12243	14786	17838	21442	25743	30812
26	10113	12265	14816	17880	21498	25820
27	8337	10124	12280	14838	17910	21540
28	6838	8342	10131	12291	14853	17932
29	5603	6840	8345	10136	12298	14864
30	*4565	*5604	6841	8347	10139	12303
31			*5604	*6842	8348	10141
32					*6842	*8349

m \ n	65	66	67	68	69	70
1	1	1	1	1	1	1
2	32	33	33	34	34	35
3	352	363	374	385	397	408
4	1991	2087	2178	2280	2376	2484
5	7166	7599	8056	8529	9027	9542
6	18467	19858	21301	22856	24473	26207
7	37108	40340	43819	47527	51508	55748
8	61575	67696	74280	81457	89162	97539
9	88252	97922	1 08527	1 20092	1 32751	1 46520
10	1 12804	1 26299	1 41136	1 57564	1 75586	1 95491
11	1 31970	1 48847	1 67672	1 88556	2 11782	2 37489
12	1 43948	1 63540	1 85425	2 10005	2 37405	2 68079
13	1 48702	1 69919	1 93906	2 20877	2 51274	2 85373
14	1 47059	1 69027	1 93880	2 22118	2 53981	2 90071
15	1 40587	1 62331	1 87175	2 15415	2 47587	2 84054
16	1 30738	1 51685	1 75618	2 03067	2 34343	2 70105
17	1 19009	1 38579	1 61144	1 87013	2 16738	2 50723
18	1 06408	1 24418	1 45149	1 69120	1 96648	2 28364
19	93854	1 10059	1 28886	1 50614	1 75767	2 04725
20	81801	96271	1 13039	1 32559	1 55112	1 81274
21	70667	83372	98222	1 15463	1 35550	1 58800
22	60547	71669	84626	99795	1 17417	1 37979
23	51577	61174	72461	85628	1 01050	1 18991
24	43668	51962	61664	73088	86420	1 02052
25	36830	43899	52259	62049	73578	87047
26	30913	36965	44075	52490	62346	73963
27	25876	30990	37066	44210	52666	62577
28	21570	25918	31046	37143	44311	52801
29	17947	21592	25948	31088	37199	44388
30	14871	17958	21607	25970	31118	37241
31	12306	14876	17965	21618	25985	31140
32	10142	12308	14879	17970	21625	25996
33	*8349	*10143	12309	14881	17973	21630
34			*10143	*12310	14882	17975
35					*12310	*14883

m \ n	71	72	73	74	75
1	1	1	1	1	1
2	35	36	36	37	37
3	420	432	444	456	469
4	2586	2700	2808	2928	3042
5	10083	10642	11229	11835	12470
6	28009	29941	31943	34085	36308
7	60289	65117	70281	75762	81612
8	1 06522	1 16263	1 26692	1 37977	1 50042
9	1 61554	1 77884	1 95666	2 14944	2 35899
10	2 17280	2 41279	2 67507	2 96320	3 27748
11	2 66006	2 97495	3 32337	3 70733	4 13112
12	3 02196	3 40293	3 82562	4 29636	4 81769
13	3 23689	3 66566	4 14624	4 68273	5 28245
14	3 30699	3 76577	4 28104	4 86133	5 51155
15	3 25472	3 72311	4 25349	4 85184	5 52767
16	3 10748	3 57075	4 09603	4 69300	5 36827
17	2 89656	3 34051	3 84759	4 42442	5 08137
18	2 64691	3 06421	3 54091	4 08687	4 70914
19	2 38134	2 76493	3 20620	3 71153	4 29112
20	2 11428	2 46288	2 86364	3 32557	3 85528
21	1 85794	2 16965	2 53033	2 94574	3 42505
22	1 61798	1 89493	2 21500	2 58592	3 01349
23	1 39935	1 64230	1 92496	2 25206	2 63138
24	1 20246	1 41510	1 66187	1 94930	2 28212
25	1 02844	1 21248	1 42765	1 67762	1 96888
26	87537	1 03471	1 22040	1 43767	1 69017
27	74260	87922	1 03961	1 22667	1 44559
28	62753	74491	88219	1 04346	1 23157
29	52902	62888	74667	88450	1 04643
30	44444	52979	62989	74802	88626
31	37271	44486	53035	63066	74903
32	31155	37293	44516	53077	63122
33	26003	31166	37308	44538	53107
34	21633	26008	31173	37319	44553
35	17976	21635	26011	31178	37326
36	*14883	*17977	21636	26013	31181
37			*17977	*21637	26014
38					*21637

m \ n	76	77	78	79	80
1	1	1	1	1	1
2	38	38	39	39	40
3	481	494	507	520	533
4	3169	3289	3422	3549	3689
5	13125	13811	14518	15257	16019
6	38677	41134	43752	46461	49342
7	87816	94425	1 01423	1 08869	1 16742
8	1 63069	1 76978	1 91964	2 07945	2 25132
9	2 58569	2 83161	3 09729	3 38484	3 69499
10	3 62198	3 99705	4 40725	4 85315	5 33975
11	4 59718	5 11045	5 67377	6 29281	6 97097
12	5 39672	6 03666	6 74585	7 52802	8 39286
13	5 95056	6 69555	7 52368	8 44504	9 46708
14	6 24188	7 05851	7 97341	8 99427	10 13531
15	6 28822	7 14504	8 10726	9 18851	10 40014
16	6 13370	6 99749	7 97402	9 07376	10 31391
17	5 82691	6 67382	7 63265	8 71908	9 94644
18	5 41971	6 22771	7 14802	8 19205	9 37815
19	4 95332	5 71069	6 57395	7 55880	8 67873
20	4 46405	5 16054	5 95872	6 86983	7 91131
21	3 97566	4 60915	5 33523	6 16825	7 12083
22	3 50757	4 07570	4 73030	5 48134	6 34429
23	3 06923	3 57554	4 15852	4 83076	5 60305
24	2 66849	3 11476	3 63139	4 22664	4 91380
25	2 30647	2 69857	3 15190	3 67697	4 28256
26	1 98463	2 32605	2 72293	3 18199	3 71413
27	1 70019	1 99718	2 34180	2 74251	3 20635
28	1 45186	1 70811	2 00720	2 35435	2 75826
29	1 23542	1 45676	1 71438	2 01512	2 36437
30	1 04874	1 23839	1 46061	1 71928	2 02139
31	88761	1 05050	1 24070	1 46358	1 72313
32	74980	88862	1 05185	1 24246	1 46589
33	63164	75036	88939	1 05286	1 24381
34	53129	63194	75078	88995	1 05363
35	44564	53144	63216	75108	89037
36	37331	44571	53155	63231	75130
37	31183	37334	44576	53162	63242
38	*26015	31184	37336	44579	53167
39		*26015	*31185	37337	44581
40				*31185	*37338

m \ n	81	82	83	84	85
1	1	1	1	1	1
2	40	41	41	42	42
3	547	560	574	588	602
4	3822	3969	4109	4263	4410
5	16814	17633	18487	19366	20282
6	52327	55491	58767	62239	65827
7	1 25104	1 33939	1 43307	1 53192	1 63662
8	2 43434	2 63081	2 83981	3 06376	3 30170
9	4 03016	4 39100	4 78025	5 19880	5 64945
10	5 86779	6 44295	7 06607	7 74345	8 47628
11	7 71464	8 52785	9 41790	10 38944	11 45078
12	9 34502	10 39543	11 54981	12 82083	14 21506
13	10 60163	11 85776	13 24916	14 78670	16 48649
14	11 40588	12 82281	14 39757	16 14987	18 09369
15	11 75862	13 27763	14 97696	16 87344	18 99041
16	11 70752	13 27547	15 03381	17 00763	19 21687
17	11 33373	12 89761	14 66126	16 64525	18 87776
18	10 72093	12 24262	13 96169	15 90544	18 09674
19	9 95318	11 39939	13 04117	14 90023	17 00603
20	9 09742	10 44986	11 98693	13 73531	15 71824
21	8 21104	9 45483	10 87482	12 49153	14 33286
22	7 33212	8 46435	9 75753	11 23608	12 92139
23	6 49117	7 50917	8 67699	10 01260	11 54109
24	5 70381	6 61330	7 65661	8 85481	10 22625
25	4 98203	5 78700	6 71428	7 77904	9 00267
26	4 32817	5 03800	5 85530	6 79758	7 88017
27	3 74423	4 36534	5 08363	5 91130	6 86593
28	3 22593	3 76859	4 39544	5 12081	5 95694
29	2 77081	3 24168	3 78817	4 41980	5 15091
30	2 37229	2 78083	3 25423	3 80392	4 43938
31	2 02629	2 37856	2 78875	3 26425	3 81647
32	1 72610	2 03014	2 38346	2 79502	3 27217
33	1 46765	1 72841	2 03311	2 38731	2 79992
34	1 24482	1 46900	1 73017	2 03542	2 39028
35	1 05419	1 24559	1 47001	1 73152	2 03718
36	89067	1 05461	1 24615	1 47078	1 73253
37	75145	89089	1 05491	1 24657	1 47134
38	63249	75156	89104	1 05513	1 24687
39	53170	63254	75163	89115	1 05528
40	44582	53172	63257	75168	89122
41	*37338	*44583	53173	63259	75171
42			*44583	*53174	63260
43					*53174

m \ n	86	87	88	89	90
1	1	1	1	1	1
2	43	43	44	44	45
3	616	631	645	660	675
4	4571	4725	4894	5055	5231
5	21224	22204	23212	24260	25337
6	69624	73551	77695	81979	86499
7	1 74696	1 86366	1 98655	2 11634	2 25286
8	3 55626	3 82641	4 11498	4 42089	4 74715
9	6 13331	6 65355	7 21125	7 80997	8 45105
10	9 27143	10 13036	11 06080	12 06440	13 14972
11	12 60740	13 86861	15 24081	16 73457	18 35721
12	15 74714	17 42509	19 26533	21 27747	23 48042
13	18 36130	20 42987	22 70754	25 21589	27 97302
14	20 25226	22 64234	25 29120	28 21909	31 45777
15	21 34841	23 97537	26 89583	30 14304	33 74676
16	21 69146	24 45589	27 54612	30 99186	34 83604
17	21 38425	24 19869	27 35245	30 88663	34 83945
18	20 56896	23 35073	26 48233	29 99936	33 95084
19	19 38560	22 07510	25 10840	28 52958	32 38070
20	17 96874	20 51599	23 40069	26 65952	30 34232
21	16 42491	18 80246	21 49821	24 55532	28 01502
22	14 84373	17 03038	19 51915	22 34447	25 55327
23	13 28562	15 27744	17 54615	20 13089	23 06908
24	11 79751	13 59239	15 64398	17 98283	20 65051
25	10 40463	12 01193	13 84982	15 95210	18 35113
26	9 12532	10 55279	12 19073	14 06480	16 21030
27	7 96354	9 22656	10 67559	12 33911	14 24390
28	6 92195	8 03192	9 30998	10 77690	12 46202
29	5 99412	6 96760	8 08795	9 37838	10 86035
30	5 17527	6 02422	7 00478	8 13360	9 43442
31	4 45513	5 19485	6 04858	7 03488	8 17078
32	3 82649	4 46768	5 21060	6 06816	7 05924
33	3 27844	3 83441	4 47770	5 22315	6 08391
34	2 80377	3 28334	3 84068	4 48562	5 23317
35	2 39259	2 80674	3 28719	3 84558	4 49189
36	2 03853	2 39435	2 80905	3 29016	3 84943
37	1 73330	2 03954	2 39570	2 81081	3 29247
38	1 47176	1 73386	2 04031	2 39671	2 81216
39	1 24709	1 47206	1 73428	2 04087	2 39748
40	1 05539	1 24724	1 47228	1 73458	2 04129
41	89127	1 05546	1 24735	1 47243	1 73480
42	75173	89130	1 05551	1 24742	1 47254
43	*63261	75174	89132	1 05554	1 24747
44		*63261	*75175	89133	1 05556
45				*75175	*89134

m \ n	91	92	93	94	95
1	1	1	1	1	1
2	45	46	46	47	47
3	690	705	721	736	752
4	5400	5584	5760	5952	6136
5	26455	27604	28796	30020	31289
6	91164	96079	1 01155	1 06491	1 11999
7	2 39691	2 54826	2 70775	2 87521	3 05146
8	5 09267	5 46067	5 84996	6 26401	6 70162
9	9 13815	9 87292	10 65947	11 49941	12 39732
10	14 31884	15 58110	16 93899	18 40292	19 97569
11	20 12069	22 03348	24 10895	26 35689	28 79236
12	25 88523	28 51355	31 37850	34 50438	37 90670
13	31 00410	34 33027	37 98063	41 98013	46 36214
14	35 03153	38 97751	43 32454	48 11594	53 38601
15	37 74599	42 17657	47 08477	52 51305	58 51608
16	39 11503	43 87969	49 17406	55 05879	61 58681
17	39 26046	44 19640	49 70660	55 84788	62 69144
18	38 38036	43 34733	48 90554	55 12631	62 07559
19	36 71577	41 58656	47 05886	53 19666	60 07963
20	34 49498	39 17865	44 45020	50 38443	57 05194
21	31 93032	36 35292	41 34830	46 98053	53 33017
22	29 18919	33 31011	37 97090	43 24323	49 19553
23	26 40955	30 19969	34 50002	39 37025	44 88553
24	23 68572	27 14043	31 06389	35 52054	40 57271
25	21 08950	24 20831	27 76092	31 79967	36 39101
26	18 66026	21 45915	24 64906	28 28582	32 42313
27	16 42570	18 91902	21 76905	25 01972	28 72792
28	14 39243	16 60502	19 13472	22 02823	25 33018
29	12 56338	14 51541	16 75366	19 31419	22 24415
30	10 92876	12 64685	14 61680	16 87669	19 46290
31	9 48007	10 98480	12 71527	14 70028	16 97810
32	8 20088	9 51725	11 03045	12 77131	14 76870
33	7 07882	8 22524	9 54735	11 06763	12 81696
34	6 09646	7 09457	8 24482	9 57171	11 09773
35	5 24109	6 10648	7 10712	8 26057	9 59129
36	4 49679	5 24736	6 11440	7 11714	8 27312
37	3 85240	4 50064	5 25226	6 12067	7 12506
38	3 29423	3 85471	4 50361	5 25611	6 12557
39	2 81317	3 29558	3 85647	4 50592	5 25908
40	2 39804	2 81394	3 29659	3 85782	4 50768
41	2 04159	2 39846	2 81450	3 29736	3 85883
42	1 73495	2 04181	2 39876	2 81492	3 29792
43	1 47261	1 73506	2 04196	2 39898	2 81522
44	1 24750	1 47266	1 73513	2 04207	2 39913
45	1 05557	1 24752	1 47269	1 73518	2 04214
46	*89134	*1 05558	1 24753	1 47271	1 73521
47			*1 05558	*1 24754	1 47272
48					*1 24754

m \ n	96	97	98	99	100
1	1	1	1	1	1
2	48	48	49	49	50
3	768	784	800	817	833
4	6336	6528	6736	6936	7153
5	32591	33940	35324	36756	38225
6	1 17788	1 23755	1 30019	1 36479	1 43247
7	3 23633	3 43074	3 63446	3 84845	4 07254
8	7 16644	7 65722	8 17789	8 72713	9 30912
9	13 35517	14 37769	15 46719	16 62894	17 86528
10	21 66875	23 48553	25 43849	27 53159	29 77866
11	31 42647	34 27615	37 35414	40 67930	44 26616
12	41 61319	45 64153	50 02329	54 77923	59 94463
13	51 15586	56 39989	62 12802	68 38459	75 20910
14	59 18495	65 55343	72 54976	80 22171	88 63685
15	65 14463	72 46258	80 53039	89 42320	99 21212
16	68 82999	76 85215	85 73805	95 56420	106 43083
17	70 30589	78 77643	88 18588	98 63566	110 22546
18	69 83946	78 49794	88 15458	98 90681	110 87828
19	67 78628	76 41341	86 05674	96 83331	108 85999
20	64 54368	72 94682	82 37213	92 92657	104 74462
21	60 47699	68 51934	77 55597	87 70736	99 09482
22	55 91609	63 49048	72 02691	81 63167	92 43766
23	51 12049	58 16815	66 12186	75 09614	85 20721
24	46 30063	52 78236	60 11745	68 40398	77 76463
25	41 60115	47 51311	54 21001	61 79507	70 37286
26	37 13064	42 47652	48 54782	55 43041	63 23274
27	32 94979	37 75641	43 21912	49 42704	56 47002
28	29 09935	33 39290	38 28442	43 84665	50 17195
29	25 58966	29 41023	33 76489	38 72830	44 37567
30	22 42373	25 80573	29 66993	34 07607	39 10071
31	19 58596	22 57249	25 98538	29 88611	34 33592
32	17 06159	19 68738	22 69557	26 13417	30 06581
33	14 82474	17 13001	19 77087	22 79700	26 25726
34	12 85414	14 87039	17 18605	19 83929	22 88049
35	11 12209	12 88424	14 90757	17 23170	19 89533
36	9 60704	11 14167	12 90860	14 93767	17 26888
37	8 28314	9 61959	11 15742	12 92818	14 96203
38	7 13133	8 29106	9 62961	11 16997	12 94393
39	6 12942	7 13623	8 29733	9 63753	11 17999
40	5 26139	6 13239	7 14008	8 30223	9 64380
41	4 50903	5 26315	6 13470	7 14305	8 30608
42	3 85960	4 51004	5 26450	6 13646	7 14536
43	3 29834	3 86016	4 51081	5 26551	6 13781
44	2 81544	3 29864	3 86058	4 51137	5 26628
45	2 39924	2 81559	3 29886	3 86088	4 51179
46	2 04219	2 39931	2 81570	3 29901	3 86110
47	1 73523	2 04222	2 39936	2 81577	3 29912
48	*1 47273	1 73524	2 04224	2 39939	2 81582
49		*1 47273	*1 73525	2 04225	2 39941
50				*1 73525	*2 04226

ON THE GROUPS $\pi_r(V_{n,m})$ AND SPHERE-BUNDLES

By J. H. C. WHITEHEAD.

[Received 12 June, 1941.—Read 16 December, 1943.]

1. Let $S^\nu(K)$ be a sphere-bundle† over a simplicial complex K. The ν-sphere $S^\nu(p)$, associated with any point $p \in K$, is to have the metric of a hypersphere with unit radius in Euclidean metric space R^n $(n = \nu+1)$. Two points on $S^\nu(p)$, distant $\frac{1}{2}\pi$ apart in $S^\nu(p)$, will be described as *mutually orthogonal*. By a *normal m-fold projection* $(m \leqslant n)$ in $S^\nu(K)$, of a sub-set $F \subset K$, we shall mean an m-valued transformation of F into a sub-set of $S^\nu(K)$, in which a given point $p \in F$ corresponds to an ordered set of m mutually orthogonal points, $x_1(p), \ldots, x_m(p) \subset S^\nu(K)$, each of which varies continuously with p. The totality of normal m-fold projections in $S^\nu(K)$, of a single point $p \in K$, may obviously be identified with the space‡ $V_{n,m}$, which consists of all the ordered sets of m unit, mutually orthogonal vectors at a given point $O \in R^n$. If $m \leqslant n-2$, then§ $\pi_1(V_{n,m}) = 1$. On the other hand, $V_{n,n-1}$ is equivalent to the group of rotations in R^n, and is therefore r-simple‖, for each value of $r = 1, 2, \ldots$. It follows that an element in the homotopy group $\pi_r(V_{n,m})$ $(m < n)$ is uniquely determined by a map $f(S^r) \subset V_{n,m}$, without reference to a base point. We now assume $S^\nu(K)$ to be orientable. That is to say, there is a normal n-fold projection in $S^\nu(K)$ of each closed simplex in K such that, if p is any

† Hassler Whitney, *Bulletin American Math. Soc.*, 43 (1937), 785–805, and *Proc. Nat. Academy of Sciences*, 21 (1935), 464–468; 26 (1940), 148–153.

‡ Cf. E. Stiefel, *Comm. Math. Helvetici*, 8 (1936), 305–353.

§ Stiefel, *loc. cit.*, or § 2 below.

‖ J. H. C. Whitehead, "On the homotopy types of manifolds", *Annals of Math.*, 41 (1940), 825–832. An arcwise connected topological space X is said to be 1-simple if $\pi_1(X)$ is Abelian and r-simple $(r > 1)$ if any map $f(S^r) \subset \tilde{X}$ is homotopic to $tf(S^r) \subset \tilde{X}$, where $\tilde{X}$ is the universal covering space of X and t is any transformation in the covering group (S. Eilenberg, *Fund. Math.*, 32 (1939), 167–175). Otherwise stated, each element in $\pi_r(X)$ $(r > 1)$ is invariant under the operators of the group ring of $\pi_1(X)$ (J. H. C. Whitehead, *Proc. London Math. Soc.* (2), 45 (1939), 243–327, §§ 10, 11).

point in a boundary simplex, A, of a given simplex, $B \subset K$, then the projections of A and B determine the same orientation of $S^r(p)$. After these preliminaries we start from the following theorem, whose proof is essentially the same as one which I have given elsewhere† for the special case $m = n-1$.

THEOREM 1. *A normal m-fold projection of‡ K^r in*

$$S^r(K) \quad [0 \leqslant r < \dim(K)]$$

determines an $(r+1)$-dimensional co-cycle C^{r+1} in K^{r+1}, with coefficients in $\pi_r(V_{n,m})$. There is a normal m-fold projection of K^{r+1} in $S^r(K)$ if, and only if, $C^{r+1} \backsim 0$.

The content of this theorem will increase progressively as the groups $\pi_r(V_{n,m})$ $(m < n)$ are calculated. Stiefel§, though he did not use the idiom of homotopy groups, proved that

$$\begin{aligned} \pi_r(V_{n,m}) &= 0 \quad \text{if} \quad r < n-m \\ &= Z_2 \quad \text{if} \quad r = n-m \equiv 1 \pmod 2 \\ &= Z_\infty \quad \text{if} \quad r = n-m \equiv 0 \pmod 2, \end{aligned}$$

where Z_h stands for a cyclic group of order h $(h = 1, 2, \ldots, \infty)$. The main object of this paper is to calculate $\pi_{n-m+1}(V_{n,m})$ and $\pi_{n-m+2}(V_{n,m})$.

2. In this section we give an explicit description of $V_{n,m}$ $(m < n)$ as a sum of non-singular (open) cells, with singular boundaries. We take the points in $V_{n,m}$ to be ordered sets of mutually orthogonal points $x_1, \ldots, x_m$, on an $(n-1)$-sphere $S^{n-1} \subset R^n$. Let P^{n-1} be the projective $(n-1)$-space whose points are diametrically opposite point-pairs

$$x, x^* \subset S^{n-1},$$

and let $u(S^{n-1}) = P^{n-1}$ be the map in which $x \in S^{n-1}$ corresponds to $(x, x^*) \in P^{n-1}$. If $x, y \subset S^{n-1}$ are mutually orthogonal we also describe $u(x)$, y as orthogonal. Given $x, y \subset S^{n-1}$, we use‖ $x.y$ to stand for the mirror image of y in the hyperplane through O, which is orthogonal to Ox, where O is the centre of S^{n-1}. Clearly $x.y = x^*.y$.

† "On homotopy types of manifolds" (*loc. cit.*).

‡ K^r stands for the aggregate of simplexes in K, whose dimensionalities do not exceed r.

§ *loc. cit.*

‖ Cf. H. Hopf, *Fund. Math.*, 25 (1935), 427–440.

Therefore a single-valued product $\xi . y \in S^{n-1}$, of given points† $\xi = u(x) \in P^{n-1}$ and $y \in S^{n-1}$, is defined by $\xi . y = x . y$. The multiplication $\xi . a$ is not associative‡, and $\xi_p . \xi_{p-1} . \ldots . \xi_1 . a$ $(p > 1)$ will stand for the product which is defined inductively as§

$$\xi_p . (\xi_{p-1} . \ldots . \xi_1 . a).$$

If $a \neq b$ $(a, b \subset S^{n-1})$ the equation $\xi . a = b$ has a unique solution $\xi \in P^{n-1}$, namely $\xi = (x, x^*)$, where the diameter xx^* is parallel to the chord ab. On the other hand, $\xi . a = a$ if, and only if, $\xi \in P^{n-2}$, where P^{n-2} is the locus of points in P^{n-1} which are orthogonal to a. It is obvious that the transformation $\xi \to y = \xi . a$ is continuous, also that its inverse is continuous when $y \neq a$. Therefore $\xi \to \xi . a$ is a map of P^{n-1} on S^{n-1} in which P^{n-2} corresponds to a and $P^{n-1} - P^{n-2}$ is mapped topologically on $S^{n-1} - a$.

Let $(a_1, \ldots, a_m)$ be a fixed point in $V_{n,m}$ $(a_j \in S^{n-1})$, let S^{k+j-1} $(k = n-m)$ be the sphere consisting of the points in S^{n-1} which are orthogonal to each of $a_{j+1}, \ldots, a_m$ $(j = 0, \ldots, m-1)$, and let $P^{k+j-1} = u(S^{k+j-1})$. Then $S^{k-1} \subset S^k \subset \ldots \subset S^{n-1}$, $a_j \in S^{k+j-1}$ $(j = 1, \ldots, m)$ and $\xi_i . a_j = a_j$ if $\xi_i \in P^{k+i-1}$ and $j > i$. I say that, given $(v_1, \ldots, v_m) \in V_{n,m}$, the equations

$$v_j = \xi_m . \ldots . \xi_j . a_j \quad (j = 1, \ldots, m) \tag{2.1}$$

have solutions $\xi_i \in P^{k+i-1}$. For $\xi_m . a_m = v_m$ can be solved for $\xi_m \in P^{n-1}$, and the assertion is justified in the case $m = 1$. Since the transformation $x \to \xi . x$ is a reflexion, we have $\xi . \xi . x = x$. Therefore the equations which have still to be solved can be written in the form

$$\xi_{m-1} . \ldots . \xi_j . a_j = v_j' \quad (j = 1, \ldots, m-1),$$

where $v_j' = \xi_m . v_j$. Since v_i is orthogonal to v_j if $i < j$ it follows that $v_1', \ldots, v_{m-1}'$ are a mutually orthogonal set of points in P^{n-2}, and the assertion follows from induction on m. The solution of $\xi . a = b$ is unique if $b \neq a$, that is to say, if a given solution η is not orthogonal to a. Therefore, given $\eta_j \in \mathfrak{S}^{k+j-1}$, where $\mathfrak{S}^{k+j-1} = P^{k+j-1} - P^{k+j-2}$ $(j = 1, \ldots, m)$, it follows inductively, starting with $\xi_m . a_m = \eta_m . a_m$, which gives $\xi_m = \eta_m$, that the equations

$$\xi_m . \ldots . \xi_j . a_j = \eta_m . \ldots . \eta_j . a_j \quad (j = 1, \ldots, m)$$

† We denote points in S^{n-1} by small Roman letters, and points in P^{n-1} by small Greek letters.

‡ Clearly $f(b . c) = f(b) . f(c)$, where f is any (direct or opposite) Euclidean transformation. Taking $f(x) = a . x$, we have $a . (b . c) = (a . b) . (a . c)$.

§ *I.e.* $\xi_p . \xi_{p-1} . \ldots . \xi_1 . a = f_p f_{p-1} \ldots f_1(a)$, where $f_j(x) = \xi_j . x$.

have the unique solutions $\xi_j = \eta_j$. Now let $\xi_i \in P^{k+i-1}$ be given, and let us take $(\xi_1, \ldots, \xi_m)$ to be a point in the topological product

$$P^k \times \ldots \times P^{n-1}.$$

Then it is obvious that the transformation

$$\phi(P^k \times \ldots \times P^{n-1}) = V_{n,m},$$

which is given by (2.1), is continuous and has a continuous inverse near a point $\phi(\xi_1^0, \ldots, \xi_m^0)$, where $\xi_j^0 \in \mathfrak{E}^{k+j-1}$. Hence ϕ is a (continuous) map of $P^k \times \ldots \times P^{n-1}$ on $V_{n,m}$ and is topological throughout $\mathfrak{E}^k \times \ldots \times \mathfrak{E}^{n-1}$. Moreover, no point in $\mathfrak{E}^k \times \ldots \times \mathfrak{E}^{n-1}$ has the same image as a point in† $P^k \times \ldots \times P^{n-1} - \mathfrak{E}^k \times \ldots \times \mathfrak{E}^{n-1}$.

Let $\lambda_1, \ldots, \lambda_p$ $(1 \leqslant p \leqslant m)$ be given integers, subject only to the conditions $k-1 \leqslant \lambda_i \leqslant n-p+i-1$ $(i = 1, \ldots, p)$, and let

$$P_{k-1}^{\lambda_1 \ldots \lambda_p} = \phi(P^{k-1} \times \ldots \times P^{k-1} \times P^{\lambda_1} \times \ldots \times P^{\lambda_p}). \tag{2.2}$$

In other words, $P_{k-1}^{\lambda_1 \ldots \lambda_p}$ is the set of points $\phi(\xi_1', \ldots, \xi_{m-p}', \xi_1, \ldots, \xi_p)$, where $\xi_i' \in P^{k-1}$ and $\xi_j \in P^{\lambda_j}$. Since $\xi . a_j = a_j$ $(j = 1, \ldots, m)$ if $\xi \in P^{k-1}$, it follows from (2.1) that $P_{k-1}^{\lambda_1 \ldots \lambda_p}$ is unaltered if we replace each of the first $m-p$ factors P^{k-1}, on the right-hand side of (2.2), by an arbitrary point $\xi_0 \in P^{k-1}$. We take

$$P_{k-1}^0 = \phi(\xi_0, \ldots, \xi_0) = (a_1, \ldots, a_m)$$

to be the left-hand side of (2.2) when $p = 0$. Clearly

$$P_{k-1}^{k\, k+1 \ldots n-t-1} = V_{n-t,m-t} \quad (0 \leqslant t \leqslant m-1),$$

where $V_{n-t,m-t} \subset V_{n,m}$ $(t > 0)$ consists of the points $(v_1, \ldots, v_m)$ for which $v_{m-t+1} = a_{m-t+1}, \ldots, v_m = a_m$. Notice also that

$$P_{k-1}^{\lambda_1 \ldots \lambda_p} \subset P_{k-1}^{\mu_1 \ldots \mu_p}$$

if $\lambda_i \leqslant \mu_i$ $(i = 1, \ldots, p)$.

LEMMA 1. (i) $\quad P_{k-1}^{\lambda_1 \ldots \lambda_i \lambda_{i+1} \ldots \lambda_p} = P_{k-1}^{\lambda_1 \ldots \lambda_{i+1} \lambda_i \ldots \lambda_p}$,

provided that $\lambda_{i+1} \leqslant n-p+i-1$,

(ii) $\quad P_{k-1}^{\lambda_1 \ldots \mu \mu \ldots \lambda_p} = P_{k-1}^{\lambda_1 \ldots \mu-1 \mu \ldots \lambda_p}$.

Let ξ, η be distinct points in P^{n-1}. The transformation $p \to \eta . \xi . p$ $(p \in S^{n-1})$ is the resultant of successive reflexions about distinct hyperplanes through O, and is therefore a rotation about the $(n-2)$-space,

† This also follows from the fact that $V_{n,m}$ is a manifold.

R^{n-2}, in which they intersect. Hence $\eta \,.\, \xi \,.\, p = \eta' \,.\, \xi' \,.\, p$, where ξ', η' are the images of ξ, η in any rotation about† R^{n-2}. Therefore we may take either ξ' or η' to be a given point on the line $\xi\eta \subset P^{n-1}$, both ξ' and η' being on this line. Hence

(i) if $\xi, \eta \subset P^t \subset P^{n-1}$ there are points $\xi', \eta' \subset P^t$ such that

$$\eta \,.\, \xi \,.\, p = \xi \,.\, \eta' \,.\, p = \xi' \,.\, \eta \,.\, p,$$

for every $p \in S^{n-1}$;

(ii) if $\xi, \eta \subset P^t$, and $P^{t-1} \subset P^t$ is given, then there are points $\xi' \in P^{t-1}$, $\eta' \in P^t$, such that $\eta \,.\, \xi \,.\, p = \eta' \,.\, \xi' \,.\, p$, for every $p \in S^{n-1}$.

If $\xi = \eta$ then $\eta \,.\, \xi \,.\, p = p = \xi' \,.\, \xi' \,.\, p$, for any $\xi' \in P^{n-1}$, and these assertions are still true.

If $\eta_i \in P^{k+i-1}$ we have $\eta_i \,.\, a_j = a_j$ when $j > i$. Therefore the equations (2.1) may be written in the form

$$v_j = \eta_m \,.\, \eta_{m-1} \,.\, \dots \,.\, \eta_1 \,.\, a_j,$$

and every point $(v_1, \dots, v_m) \in P_{k-1}^{\lambda_1 \dots \lambda_p}$ is given by

$$v_j = \xi_p \,.\, \dots \,.\, \xi_1 \,.\, a_j \quad (j = 1, \dots, m),$$

for suitable values of $\xi_j \in P^{\lambda_j}$. It follows from the result (i) above that

$$\xi_{i+1} \,.\, \xi_i \,.\, \dots \,.\, \xi_1 \,.\, a_j = \xi_i' \,.\, \xi_{i+1}' \,.\, \dots \,.\, \xi_1 \,.\, a_j,$$

where we may take $\xi_i' = \xi_i$, $\xi_{i+1}' \in P^{\lambda_{i+1}}$, as we do if $\lambda_i \leqslant \lambda_{i+1}$, or $\xi_{i+1}' = \xi_{i+1}$, $\xi_i' \in P^{\lambda_i}$, as we do if $\lambda_{i+1} < \lambda_i$. Therefore

$$\xi_m \,.\, \dots \,.\, \xi_{i+1} \,.\, \xi_i \,.\, \dots \,.\, \xi_1 \,.\, a_j = \xi_m \,.\, \dots \,.\, \xi_i' \,.\, \xi_{i+1}' \,.\, \dots \,.\, \xi_1 \,.\, a_j.$$

If $\lambda_{i+1} \leqslant n-p+i-1$, the right-hand side is the component v_j of a point $(v_1, \dots, v_m) \in P_{k-1}^{\lambda_1 \dots \lambda_{i+1} \lambda_i \dots \lambda_p}$, whence

$$P_{k-1}^{\lambda_1 \dots \lambda_i \lambda_{i+1} \dots \lambda_p} \subset P_{k-1}^{\lambda_1 \dots \lambda_{i+1} \lambda_i \dots \lambda_p}.$$

Similarly $$P_{k-1}^{\lambda_1 \dots \lambda_{i+1} \lambda_i \dots \lambda_p} \subset P_{k-1}^{\lambda_1 \dots \lambda_i \lambda_{i+1} \dots \lambda_p},$$

and the first part of the lemma is established. It follows by a similar argument, from the result (ii) above, that

$$P_{k-1}^{\dots \mu \mu \dots} \subset P_{k-1}^{\dots \mu-1 \mu \dots}.$$

Clearly $$P_{k-1}^{\dots \mu-1 \mu \dots} \subset P_{k-1}^{\dots \mu \mu \dots},$$

and the lemma follows.

† *I.e.* the transformation of P^{n-1} which is determined by a rotation of S^{n-1}.

Notice that, since $\lambda_i \leqslant n-p+i-1$, the condition $\lambda_{i+1} \leqslant n-p+i-1$ is certainly satisfied if $\lambda_{i+1} < \lambda_i$.

Let†

$$\mathfrak{E}^{\lambda_1 \dots \lambda_p} = \phi(\xi_0 \times \dots \times \xi_0 \times \mathfrak{E}^{\lambda_1} \times \dots \times \mathfrak{E}^{\lambda_p}), \tag{2.3}$$

where $\xi_0 \in P^{k-1}$, and $\mathfrak{E}^{\lambda_1 \dots \lambda_p} = \mathfrak{E}^0 = P^0_{k-1}$ if $p = 0$. If $k \leqslant \lambda_1 < \lambda_2 < \dots < \lambda_p$ and if $\eta_j \in \mathfrak{E}^{\lambda_j}$ $(j = 1, \dots, p)$ are given, it follows from a previous argument that the equations

$$\xi_p \cdot \dots \cdot \xi_j \cdot a_j = \eta_p \cdot \dots \cdot \eta_j \cdot a_j$$

have unique solutions $\xi_j \in P^{\lambda_j}$, namely $\xi_j = \eta_j$. Therefore the equations (2.3) represent a topological map of the cell $\xi_0 \times \dots \times \xi_0 \times \mathfrak{E}^{\lambda_1} \times \dots \times \mathfrak{E}^{\lambda_p}$ on $\mathfrak{E}^{\lambda_1 \dots \lambda_p}$, and no point in this cell has the same image as a point on its frontier. A similar argument shows that $\mathfrak{E}^{\lambda_1 \dots \lambda_p}$ has no point in common with $\mathfrak{E}^{\mu_1 \dots \mu_q}$ if $k \leqslant \mu_1 < \dots < \mu_q$, unless $p = q$, $\lambda_i = \mu_i$. Let $v \in P^{\lambda_1 \dots \lambda_p}_{k-1}$, where v is a given point in $V_{n,m}$ and $\lambda_1 + \dots + \lambda_p$ has its minimum value for all the sets $P^{\mu_1 \dots \mu_q}_{k-1}$ which contain v. If v were not in $\mathfrak{E}^{\lambda_1 \dots \lambda_p}$ it would be in at least one of the sets $P^{\lambda_1 \dots \lambda_{i-1} \lambda_{i+1} \dots \lambda_p}$, whereas $\lambda_1 + \dots + \lambda_p$ is a minimum for the sets which contain v. If $\lambda_i = \lambda_j$, where $i < j$, it would follow from Lemma 1 (i) that λ_j could be moved back until it was next to λ_t, for some $t \geqslant i$ such that $\lambda_t = \lambda_j$ $(t < j)$. By Lemma 1 (ii) $\lambda_1 + \dots + \lambda_p$ could then be reduced. Similarly $\lambda_1 + \dots + \lambda_p$ could be reduced if $\lambda_i = k-1$ for some i. Therefore $v \in \mathfrak{E}^{\lambda_1 \dots \lambda_p}$ and $\lambda_1, \dots, \lambda_p$ are all distinct and greater than $k-1$. It follows from Lemma 1 (i) that we may rearrange $\lambda_1, \dots, \lambda_p$ in ascending order of magnitude without altering $\mathfrak{E}^{\lambda_1 \dots \lambda_p}$. Therefore we assume that either $p = 0$ or $k \leqslant \lambda_1 < \dots < \lambda_p$. Thus each point in v is in one, and only one, of the cells $\mathfrak{E}^0$, $\mathfrak{E}^{\lambda_1 \dots \lambda_p}$ $(k \leqslant \lambda_1 < \dots < \lambda_p)$. If $k \leqslant \lambda_1 < \dots < \lambda_p$, it is clear that $\dim(\mathfrak{E}^{\lambda_1 \dots \lambda_p}) = \lambda_1 + \dots + \lambda_p$, and we denote the aggregate of these cells by $\mathfrak{K}^N$, where

$$N = k + k + 1 + \dots + n - 1 = m(2n-m-1)/2.$$

Let us regard P^λ as the cell $\mathfrak{E}^\lambda$, together with its singular boundary $P^{\lambda-1}$. It follows from the rule giving the boundary of a topological product that the boundary $F(\mathfrak{E}^{\lambda_1 \dots \lambda_p})$ of $\mathfrak{E}^{\lambda_1 \dots \lambda_p}$, regarded as the image under ϕ of a singular $(\lambda_1 + \dots + \lambda_p - 1)$-sphere, covers the set

$$\sum_{i=1}^{p} P^{\lambda_1 \dots \lambda_{i-1} \lambda_{i+1} \lambda_p}_{k-1}.$$

† If $p = 1$ there is an ambiguity in the notation, which will, however, cause no confusion. In fact we may identify $\mathfrak{E}^\lambda = P^\lambda - P^{\lambda-1}$ with the cell represented by the left-hand side of (2.3), with $p = 1$.

It follows from Lemma 1 that the latter can be expressed in the form

$$\Sigma P_{k-1}^{\mu_1 \dots \mu_q} \quad (q = 0 \text{ or } k \leqslant \mu_1 < \dots < \mu_q),$$

summed over certain values of $\mu_1, \dots, \mu_q$ for which $\mu_1 + \dots + \mu_q < \lambda_1 + \dots + \lambda_p$. If we begin with

$$F(\mathcal{E}^k) = \phi(\xi_0 \times \dots \times \xi_0 \times P^{k-1})$$

$$= \xi^0,$$

it follows from induction on $\lambda_1 + \dots + \lambda_p$ that $F(\mathcal{E}^{\lambda_1 \dots \lambda_p})$ covers an exact sum of cells in $\mathfrak{K}^N$, whose dimensionalities are less than $\lambda_1 + \dots + \lambda_p$.

We now treat $\mathfrak{K}^N$ as a complex, of which $\mathcal{E}^{\lambda_1 \dots \lambda_p}$ are the cells. The closed cells of $\mathfrak{K}^N$ (*i.e.* cells together with their boundaries) are the sets $P_{k-1}^{\lambda_1 \dots \lambda_p}$, including P_{k-1}^0. By a sub-complex $\mathfrak{K}_0 \subset \mathfrak{K}^N$ we mean a sub-set of its closed cells. By $\mathfrak{K}^r$ we mean the aggregate of cells in $\mathfrak{K}^N$ whose dimensionalities do not exceed r. Notice that $\dim(\mathfrak{K}^r)$ may be less than r. For example, $\dim(\mathfrak{K}^r) = 0$ if $r < k$. If

$$\mathcal{E}^{\lambda_1 \dots \lambda_p} \quad (k \leqslant \lambda_1 < \dots < \lambda_p)$$

is a principal† cell of $\mathfrak{K}_0 \subset \mathfrak{K}^N$, and if $r+1 < \lambda_1 + \dots + \lambda_p$, it is obvious that

$$\pi_r(\mathfrak{K}_0) = \pi_r(\mathfrak{K}_0 - \mathcal{E}^{\lambda_1 \dots \lambda_p}).$$

It follows inductively that

$$\pi_r(V_{n,m}) = \pi_r(\mathfrak{K}_1),$$

where $\mathfrak{K}_1$ is any sub-complex of $\mathfrak{K}^N$ which contains $\mathfrak{K}^{r+1}$. Thus we have two useful theorems:

THEOREM 2. *If* $k \leqslant r \leqslant n-1$, *then*

$$\pi_r(V_{n,m}) = \pi_r(V_{r+2,\,r-k+2}) = \pi_r(V_{r+1,\,r-k+1} + \mathcal{E}^{r+1}),$$

and if $k > 1$, $m > 1$, *then*

$$\pi_{n-1}(V_{n,m}) = \pi_{n-1}(V_{n-1,\,m-1} + \mathcal{E}^{n-1}).$$

For

$$V_{s,\,s-k} = P_{k-1}^{k \dots s-1} \quad (s > k),$$

whence $\mathfrak{K}^{r+1} \subset V_{r+2,\,r-k+2}$, and $V_{r+1,\,r-k+1}$ contains every cell in $\mathfrak{K}^{r+1}$ except $\mathcal{E}^{r+1}$. Also $\lambda_1 + \dots + \lambda_p > n$ if $p > 1$, $\lambda_p > n-2$, $\lambda_1 \geqslant k > 1$. Therefore, if $k > 1$, $V_{n-1,\,m-1}$ contains every cell in $\mathfrak{K}^n$ except $\mathcal{E}^{n-1}$.

† *I.e.* one which is not on the boundary of any other cell in $\mathfrak{K}_0$.

THEOREM 3. *If* $r < 2k$, *then*

$$\pi_r(V_{n,m}) = \pi_r(P^l_{k-1}),$$

where $l = \min(r+1,\ n-1)$.

If $k \leqslant \lambda_1 < \ldots < \lambda_p$ and $p > 1$, then $\lambda_1 + \ldots + \lambda_p \geqslant 2k+1 > r+1$. Therefore

$$\mathfrak{R}^{r+1} = P^{n-1}_{k-1} \quad \text{if} \quad r+1 \geqslant n-1,$$

$$= P^{r+1}_{k-1} \quad \text{if} \quad r+1 \leqslant n-1,$$

and the theorem follows.

The space P^r_{k-1} $(r \geqslant k)$ may be obtained from P^r by shrinking P^{k-1} into a point. For the map $\phi(P^r) = P^r_{k-1}$ is given by the equations $v_j = \xi . a_j$ $(j = 1, \ldots, m;\ \xi \in P^r)$. If $(v_1, \ldots, v_m) \in P^r_{k-1} - (a_1, \ldots, a_m)$ we have $v_j \neq a_j$ for at least one value of j. Therefore the equations $\xi . a_j = v_j$ have but one solution $\xi \in P^r - P^{k-1}$, whence ϕ maps $P^r - P^{k-1}$ topologically on

$$P^r_{k-1} - (a_1, \ldots, a_m).$$

On the other hand, $\phi(P^{k-1}) = (a_1, \ldots, a_m)$, since $\xi . a_j = a_j$ if $\xi \in P^{k-1}$. In particular P^k_{k-1} is the k-sphere $V_{k+1,1}$. In general

$$P^r_{k-1} = \mathfrak{E}^0 + \mathfrak{E}^k + \ldots + \mathfrak{E}^r,$$

where $F(\mathfrak{E}^k) = \mathfrak{E}^0$, $F(\mathfrak{E}^\lambda) = P^{\lambda-1}_{k-1}$ $(\lambda = k+1, \ldots, r)$.

Let S^{n+1} be given by

$$x_1^2 + \ldots + x_n^2 = 1$$

in rectangular Cartesian coordinates $x_1, \ldots, x_n$. Using the summation convention, we can easily verify that the coordinates of $x . y$ $(x, y \subset S^{n-1})$ are

$$y_i - 2x_i x_j y_j.$$

Taking a_ρ to be the point whose coordinates are $\delta_{k+\rho,i}$ $(\rho = 1, \ldots, m)$, we find that P^{n-1}_{k-1} is given explicitly by the equations

$$v_{\rho,i} = \delta_{k+\rho,i} - 2x_i x_{k+\rho}. \tag{2.4}$$

THEOREM 5. P^r_{k-1} *can be triangulated.*

Let $S^{r-k}_0 \subset S^r$ be the sphere which is orthogonal to S^{k-1} and let $P^{r-k}_0 = u(S^{r-k}_0)$. Let Π^{k-1}, Π^{r-k}_0 be triangulations of P^{k-1}, P^{r-k}_0, which we may assume to be rectilinear in terms of the projective geometry of P^r. Since a simplex is simply connected, it follows that $\Sigma^{k-1} = u^{-1}(\Pi^{k-1})$ and $\Sigma^{r-k}_0 = u^{-1}(\Pi^{r-k}_0)$ are triangulations of S^{k-1} and $S_0{}^{-k}$, and they are clearly

invariant with respect to the reflexion in the centre of S^r. The join† $\Sigma^r = \Sigma^{k-1}\Sigma_0^{r-k}$ is a triangulation of S^r such that $\Pi^r = u(\Sigma^r)$ is a rectilinear triangulation of P^r, and u is a simplicial map of Σ^r on Π^r. Let us star each simplex $AB \subset \Pi^r$, where $A \subset \Pi^{k-1}$, $B \subset \Pi_0^{r-k}$, neither A nor B being the "empty" simplex, to form a sub-division $\sigma\Pi^r$ of Π^r. Let K_1^r be the simplicial neighbourhood of Π_0^{r-k} in $\sigma\Pi^r$ and let K_1^{r-1} be the (mod 2) boundary of K_1^r. Each point in $P^r - (P^{k-1} + P_0^{r-k})$ is on one, and only one, line $\xi\xi'$, where $\xi \in P^{k-1}$, $\xi' \in P_0^{r-k}$. Therefore the neighbourhood of Π^{k-1}, which is bounded by K_1^{r-1}, is filled by a family of linear segments of the form $\xi\eta$, where $\xi \in \Pi^{k-1}$, $\eta \in K_1^{r-1}$, no two of which meet each other, except possibly at a common end point in Π^{k-1}. It is obvious that $\xi\eta$ varies continuously with η. Therefore P_{k-1}^r is covered by the simplicial complex

$$L^r = \mathfrak{E}^0 K^{r-1} + K^r,$$

where $K^{r-1} = \phi(K_1^{r-1})$, $K^r = \phi(K_1^r)$, $\mathfrak{E}^0\phi(\eta) = \phi(\xi\eta)$, and the theorem is established.

Let $\sigma\Sigma^r$ be the sub-division $u^{-1}(\sigma\Pi^r)$, of Σ^r, and let $K_0^{r-1} = u^{-1}(K_1^{r-1})$, $K_0^r = u^{-1}(K_1^r)$. Then the map $\phi u(K_0^r) = K^r$ is obviously simplicial. Let $\psi(\sigma\Sigma^r) = L^r$ be the simplicial map which is defined by $\psi = \phi u$ in K_0^r, $\psi(AB) = \mathfrak{E}^0\phi u(B)$ for given simplexes‡ $A \subset \Sigma^{k-1}$, $B \subset K_0^{r-1}$.

LEMMA 2. *The map $\phi u(S^r) \subset P_{k-1}^r$ is homotopic to $\psi(S^r) \subset P_{k-1}^r$.*

Let X^t be the sub-complex of $\sigma\Sigma^r$ which consists of K_0^r, together with all the remaining simplexes whose dimensionalities do not exceed t ($t = -1, 0, \ldots, r$; $X^{-1} = K_0^r$). Assume that ϕu has been deformed into a map $\psi_t(S^r) \subset P_{k-1}^r$, which coincides with ψ in X^t and is such that§ $\psi_t(S^r - |K_0^r|) \subset |\mathfrak{E}^0 K^{r-1}|$, for a given value of $t < r$. Then

$$\psi(A^{t+1}) + \psi_t(A^{t+1}) \subset |\mathfrak{E}^0 K^{r-1}|$$

may be regarded as a (possibly singular) topological $(t+1)$-sphere, where A^{t+1} is any $(t+1)$-simplex in $\sigma\Sigma^r$, whose interior lies in $S^r - |K_0^r|$. Since any map $f(S^{t+1}) \subset |\mathfrak{E}^0 K^{r-1}|$ is inessential‖ in $|\mathfrak{E}^0 K^{r-1}|$ it follows¶ that ψ_t

† A simplex in $\mathfrak{Z}^r$, which is not in $\mathfrak{Z}^{k-1}$ or in $\mathfrak{Z}_0^{r-k}$, is swept out by the shorter arc xy of the great circle through variable points $x \in A$, $y \in B$, where A, B are simplexes in $\mathfrak{Z}^{k-1}$, $\mathfrak{Z}_0^{r-k}$.

‡ Here B, but not A, may be the empty simplex, in which case $AB = A$.

§ $|K|$ stands for the set of points covered by a given complex K.

‖ By inessential we mean deformable into a point.

¶ Cf. W. Hurewicz, *Proc. Akad. Amsterdam*, 39 (1935), 215–224, in particular pp. 217–218.

can be deformed into a similar map ψ_{t+1}. Therefore, starting with $\psi_{-1} = \phi u$, the lemma follows from induction on t.

Let S^r be given by $x_0^2+\ldots+x_r^2 = 1$ in rectangular Cartesian coordinates for R^{r+1}, and let S^{k-1} and S^{r-k} be given by this equation, together with $x_k = \ldots = x_r = 0$ and $x_0 = \ldots = x_{k-1} = 0$, respectively. Let $M_0^{r-1} \subset S^r$ be the manifold which is given by

$$x_0^2+\ldots+x_{k-1} = x_k^2+\ldots+x_r^2 = \tfrac{1}{2}.$$

Then $M^{r-1} = u(M_0^{r-1})$ is a quadric in P^r, and P^r_{k-1} may be represented by the "truncated cone" $\mathfrak{E}^0\phi(M^{r-1})$, together with a certain one of the regions into which P^r is separated by M^{r-1}. We can also represent† P^r_{k-1} as the algebraic locus in P^N $[N = (r+1)(r-k+1)]$, which is given by

$$X_0 = x_0^2+\ldots+x_r^2, \quad X_{i\lambda} = x_i x_\lambda \quad (i = 0, \ldots, r;\ \lambda = k, \ldots, r),$$

where X_0, $X_{i\lambda}$ are homogeneous coordinates for P^N.

3. Let $\alpha \in \pi_r(P^r_{k-1})$ $(1 \leqslant k \leqslant r)$ be the element which is represented by the map‡ $\phi u(S^r) \subset P^r_{k-1}$.

THEOREM§ 5. *If $r = k$ and is even, then $\alpha = 0$; otherwise $\alpha \neq 0$. If r is odd or if $k = 1$, then α is of infinite order. If r is even and $k > 1$, then $2\alpha = 0$.*

If $r = k$, then $P^r_{k-1} = S^r$ and ϕu has degree $1-(-1)^r$. If r is even it follows that ϕu is inessential, whence $\alpha = 0$. If r is odd, then P^r is orientable and the simplexes in K^r and K^{r-1} may be so oriented that $-K^{r-1}$ is the chain boundary of K^r, where K^r and K^{r-1} mean the same as in Theorem 4. When K^r is thus oriented L^r is a cycle. The map ϕ obviously has degree ± 1, and u has degree ± 2, whence ϕu has degree ± 2. Therefore a map which represents the element $t\alpha$ has degree $\pm 2t$, and is essential unless $t = 0$. Therefore $t\alpha \neq 0$ unless $t = 0$.

If $k = 1$, then P^r_{k-1} is a projective r-space, $\phi(P^r) = P^r_{k-1}$ is a homeomorphism and $\phi u(S^r) = P^r_{k-1}$ is locally (1.1). Therefore α generates the infinite cyclic group $\pi_r(P^r_{k-1})$ and is of infinite order.

---

† This representation of P^r_{k-1} was pointed out to me by W. V. D. Hodge. Of course it follows at once from this representation that P^r_{k-1} can be triangulated (see B. L. van der Waerden, *Math. Annalen*, 102 (1929), 337–362, Appendix 1).

‡ Here, and in similar statements, it is to be understood that an orientation is assigned to S^r.

§ (*Added in proof, June*, 1943.) When writing this paper I was unable to discover whether or no $\phi u(S^4) \subset P_2^4$ was essential, but was able to show that it is essential if $\pi_4(V_{6,5}) = 0$ (see Theorem 12 *et seq.* in §6 below). B. Eckmann [*Comm. Helvetici*, 14 (1941–2), 234–256; Theorem VII, 2] has since proved that $\pi_4(V_{6,5}) = 0$.

We next take r to be even, $k > 1$, and show that $2a = 0$. Since K^{r-1} is connected, it is easily verified that any circuit in P^r_{k-1} can be deformed into K^r, and is therefore of the form $\phi(\Sigma^1)$, where Σ^1 is a circuit in P^r. But Σ^1 is homotopic in P^r to a multiple of a given projective line P^1, and since $k > 1$ we may take $P^1 \subset P^{k-1}$. Therefore $\phi(\Sigma^1)$ is homotopic to a point, namely $\mathfrak{E}^0$. Therefore P^r_{k-1} is simply connected, and hence r-simple. Thus a unique element of $\pi_r(P^r_{k-1})$ is represented by a free map $f(\Sigma^r) \subset P^r_{k-1}$, where Σ^r is an oriented topological r-sphere.

Let S^r be given by $x_0^2+\ldots+x_r^2 = 1$, and let E_0^r be the cell which is given by

$$x_r = 0, \quad x_0^2+\ldots+x_r^2 \leqslant 1.$$

Let $E_1^r \subset S^r$ be the hemisphere in which $x_r \geqslant 0$ and $E_2^r \subset S^r$ the hemisphere in which $x_r \leqslant 0$, and let $g_\rho(E_\rho^r) = E_0^r$ $(\rho = 1, 2)$ be the projection of E_ρ^r on E_0^r, which is given by $g_\rho(x_0, \ldots, x_{r-1}, x_r) = (x_0, \ldots, x_{r-1})$. Let S^r be oriented, let E_1^r, E_2^r be oriented so that $S^r = E_1^r + E_2^r$ and let E_0^r be oriented so that $\dot{E}_0^r = +\dot{E}_1^r = -\dot{E}_2^r$, where $\dot{E}$ denotes the boundary of E. Then $E_\rho^r + (-1)^\rho E_0^r$ $(\rho = 1, 2)$ is an oriented (topological) r-sphere, and a is represented by the map

$$\phi_\rho(E_\rho^r + (-1)^\rho E_0^r) \subset P^r_{k-1},$$

as well as by $\phi u(S^r) \subset P^r_{k-1}$, where

$$\begin{aligned} \phi_\rho &= \phi u \quad \text{in} \quad E_\rho^r \\ &= \phi u g_\sigma^{-1} \quad \text{in} \quad E_0^r \quad (\sigma \neq \rho). \end{aligned}$$

Let $\theta(R^{r+1}) = R^{r+1}$ be the transformation which is given by

$$\theta(x_0, \ldots, x_r) = (-x_0, \ldots, -x_{r-1}, x_r).$$

Since r is even θ is sense-preserving, and a is also represented by the map

$$\phi_\rho \theta(E_\rho^r + (-1)^\rho E_0^r) \subset P^r_{k-1}.$$

Since $\phi u(x_0, \ldots, x_r) = \phi u(-x_0, \ldots, -x_r)$ we have $\phi_1 = \phi_2 \theta$ in E_0^r. Let

$$Q^r = |E_0^r| + |E_1^r| + |E_2^r|,$$

and let $f_0(Q^r) \subset P^r_{k-1}$ be the map which is given by

$$\begin{aligned} f_0 &= \phi_1 \quad \text{in} \quad |E_0^r| + |E_1^r| \\ &= \phi_2 \theta \quad \text{in} \quad |E_0^r| + |E_2^r|. \end{aligned}$$

Clearly Q^r is simply connected, since $r > 1$, whence a free map, $f(\Sigma^r) \subset Q^r$, represents a unique element in $\pi_r(Q^r)$. Let $\beta_\rho \in \pi_r(Q^r)$ be the element

which is represented by the identical map of $E_\rho^r+(-1)^\rho E_0^r$ on itself. Clearly $\pi_s(Q^r)=0$ if $s<r$, and it follows from the consequent isomorphism† of $\pi_r(Q^r)$ on the r-th homology group of Q^r that $\beta_1+\beta_2$ is represented by the identical map of S^r on itself. Since the transformation θ preserves sense, β_2 is also represented by the map $\theta(E_2^r+E_0^r)\subset Q^r$.

Let h be the homomorphism of $\pi_r(Q^r)$ in $\pi_r(P_{k-1}^r)$, which is determined by the map $f_0(Q^r)\subset P_{k-1}^r$. Then $h(\beta_1)=h(\beta_2)=\alpha$ and $2\alpha=h(\beta_1+\beta_2)$, whence 2α is represented by $f_0(S^r)\subset P_{k-1}^r$. Since

$$\begin{aligned} f_0(S^r) &= \phi u(E_1^r) \quad \text{in} \quad |E_1^r|, \\ &= \phi u\theta(E_2^r) \quad \text{in} \quad |E_2^r|, \end{aligned}$$

and since $\phi u(x_0, \ldots, x_r)=\phi u(-x_0, \ldots, -x_r)$ and

$$\theta(x_0, \ldots, x_r)=(-x_0, \ldots, -x_{r-1}, x_r),$$

the map $f_0(S^r)$ is also given by

$$\begin{aligned} f_0(E_1^r) &= \phi u(E_1^r) \quad \text{in} \quad E_1^r, \\ f_0(E_2^r) &= \phi u\chi(E_1^r) \quad \text{in} \quad E_2^r, \end{aligned}$$

where $\chi(E_1^r)=E_2^r$ is given by $\chi(x_0, \ldots, x_{r-1}, x_r)=(x_0, \ldots, x_{r-1}, -x_r)$. Thus $f_0(S^r)=\phi u\chi'(S^r)$, where $\chi'(S^r)=E_1^r$ is given by

$$\begin{aligned} \chi'(x) &= x \quad \text{if} \quad x\in E_1^r \\ &= \chi(x) \quad \text{if} \quad x\in E_2^r. \end{aligned}$$

Therefore $f_0(S^r)$ is transformed into the constant map $f_1(S^r)=f_0(0, \ldots, 0, 1)$ by the deformation

$$f_t=\delta_t f_0 \quad (0\leqslant t\leqslant 1),$$

where $\delta_t(E_1^r)\subset E_1^r$ is a deformation of the identical map δ_0 into the constant map $\delta_1(E_1^r)=(0, \ldots, 0, 1)$. Therefore $2\alpha=0$ when r is even and $k>1$.

It remains to prove that $\alpha\neq 0$ if r is even and $r-k>0$, $k>1$, $P_k^r\neq P_2^4$. We first take $r\geqslant 2k-1$, so that $r-k\geqslant k-1$, and assume that $\alpha=0$. It follows from Lemma 2 that α is also represented by the simplicial map $\psi(\sigma\Sigma^r)\subset L^r$, where ψ, $\sigma\Sigma^r$ and L^r mean the same as in the paragraph containing Lemma 2. Since $\alpha=0$, by hypothesis, the map ψ can be extended to a map $\psi(E^{r+1})\subset P_{k-1}^r$, where E^{r+1} is an $(r+1)$-element which is bounded by S^r. By Theorem 15, in note 1, § 8 below, $\psi(E^{r+1})$ is homotopic to a map $\psi_1(E_1^{r+1})\subset L^r$, which coincides with ψ in S^r and is

† W. Hurewicz, *Proc. Akad. Amsterdam*, 38 (1935), 521–528, Theorem 1.

simplicial with respect to L^r and a triangulation† E_1^{r+1} of E^{r+1}, such that the complex $\dot{E}_1^{r+1}$ coincides with $\sigma\Sigma^r$. Since $2k-2<r$ there are (flat) projective spaces P_1^{k-1}, $P_2^{r-k} \subset |K_1^r| \subset P^r$, which do not meet each other‡. Without changing the notation, we assume K^r to have been so finely sub-divided initially that $\phi(P_1^{k-1})$ and $\phi(P_2^{r-k})$ can be deformed into cycles§ (mod 2) $\bar{C}_1^{k-1}$ and $\bar{C}_2^{r-k}$, which are on the dual‖ of L^r, without meeting each other or leaving $|K^r|$ during the deformation. Since $\psi_1 = \phi u$ in $|K_0^r|$, and since ϕu is locally (1, 1) in $|K_0^r|$, it follows that $u^{-1}(P_1^{k-1})$ and $u^{-1}(P_2^{r-k})$ can be deformed into the cycles $u^{-1}\phi^{-1}(\bar{C}_1^{k-1})$ and $u^{-1}\phi^{-1}(\bar{C}_2^{r-k})$ without meeting each other during the deformation. Since $u^{-1}(P_1^{k-1})$ and $u^{-1}(P_2^{r-k})$ are great spheres and link each other with coefficient 1, it follows that the cycles $u^{-1}\phi^{-1}(\bar{C}_1^{k-1})$ and $u^{-1}\phi^{-1}(\bar{C}_2^{r-k})$ link each other with coefficient 1. We shall arrive at a contradiction by proving that

$$u^{-1}\phi^{-1}(\bar{C}_i^h) \sim 0 \quad \text{in} \quad |\psi_1^{-1}(\bar{C}_i^h)| \subset E_1^{r+1} \quad (i=1,\,2;\ h=k-1,\,r-k),$$

which is absurd, since $|\psi_1^{-1}(\bar{C}_1^{k-1})|$ does not meet $|\psi_1^{-1}(\bar{C}_2^{r-k})|$. This will follow when we have proved that, if $\bar{C}^{r-p}$ is a given cycle on the dual of L^r, then $u^{-1}\phi^{-1}(\bar{C}^{r-p}) \sim 0$ in $|\psi_1^{-1}(\bar{C}^{r-p})|$.

Let $A_1^p, \ldots, A_{\alpha_p}^p$ be the p-simplexes in E_1^{r+1} $(p=0, \ldots, r+1)$ and let $\bar{A}_1^q, \ldots, \bar{A}_{\beta_q}^q$ be the q-simplexes in L^r $(q=0, \ldots, r)$. Let E' and L' be normal subdivisions of E_1^{r+1} and L^r, the new simplexes being such that ψ_1 maps E' simplicially¶ on L', and let a_i^p, $\bar{a}_j^q$ be the vertices of E' and L' which are internal to A_i^p and $\bar{A}_j^q$. Then $\psi_1(a_i^p) = \bar{a}_j^q$ if $\psi_1(A_i^p) = \bar{A}_j^q$. Let D_i^{r-p+1} and $\bar{D}_j^{r-q}$ be the dual stars of A_i^p and $\bar{A}_j^q$. That is to say,

$$D_i^{r-p+1} = \Sigma\, a_i^p\, a_{i_1}^{p+1} \ldots a_{i_{r-p+1}}^{r+1}, \quad \bar{D}^{r-q} = \Sigma\, \bar{a}_j^q\, \bar{a}_{j_1}^{q+1} \ldots \bar{a}_{j_{r-q}}^r,$$

where the summations extend over all values of $i_1, \ldots, i_{r-p+1}$ and $j_1, \ldots, j_{r-q}$ such that

$$A_i^p \subset A_{i_s}^{p+s}, \quad \bar{A}_j^q \subset \bar{A}_{j_t}^{q+t}.$$

---

† For future reference we assume, as we obviously may, that E_1^{r+1} is combinatorially equivalent to an $(r+1)$-simplex (*i.e.* it is not a triangulation of the kind which might be possible if the "Hauptvermutung" were false).

‡ *E.g.* P_1^{k-1} may be obtained by slightly displacing a $(k-1)$-space $P_0^{k-1} \subset P_0^{r-k}$, and we may take $P_2^{r-k} = P_0^{r-k}$.

§ Throughout the rest of this proof the coefficients in all chains are residue classes (mod 2), except where the contrary is indicated.

‖ Cf. the following paragraph.

¶ To avoid any questions which may be raised by this assumption, we may replace the loci $|E_1^{r+1}|$ and $|L^r|$ by rectilinear, polyhedral models of the complexes E_1^{r+1} and L^r.

I say that, if

$$\bar{C}^{r-p} = \Sigma \bar{c}_j \bar{D}_j^{r-p} \quad (0 \leqslant p \leqslant r), \tag{3.1}$$

for given values of $\bar{c}_1, \ldots, \bar{c}_{\beta_p}$, then†

$$\psi_1^{-1}(\bar{C}^{r-p}) = \Sigma c_i D_i^{r-p+1}, \tag{3.2}$$

where $c_i = \bar{c}_j$ if $\psi_1(A_i^p) = \bar{A}_j^p$ and $c_i = 0$ if $\dim \{\psi_1(A_i^p)\} < p$. For, since $\psi_1(A_i^p) \subset \psi_1(A_{i_s}^{p+s})$ if $A_i^p \subset A_{i_s}^{p+s}$, it follows that $\psi_1(D_i^{r-p+1}) \subset \bar{D}_j^{r-p}$ if $\psi_1(A_i^p) = \bar{A}_j^p$. Therefore $\psi_1(\Sigma c_i D_i^{r-p+1}) \subset \bar{C}^{r-p}$. Conversely, since

$$\psi_1(a_\lambda^s) = \bar{a}_\mu^t,$$

where $\bar{A}_\mu^t = \psi_1(A_\lambda^s)$, it follows from the definition of $\bar{D}_j^{r-p}$ that, if $a_\lambda^s \epsilon \psi_1^{-1}(\bar{D}_j^{r-p})$, then $\bar{A}_j^p \subset \psi_1(A_\lambda^s)$. Therefore there is a simplex $A_i^p \subset A_\lambda^s$ such that $\psi_1(A_i^p) = \bar{A}_j^p$, and $a_\lambda^s \epsilon D_\lambda^{r-s+1} \subset D_i^{r-p+1}$. It follows that

$$\psi_1^{-1}(\bar{C}^{r-p}) \subset \Sigma c_i D_i^{r-p+1},$$

and we have the equality (3.2).

If $\Sigma \bar{c}_j \bar{D}_j^{r-p}$ is a cycle, then $\Sigma \bar{c}_j \bar{A}_j^p$ is a co-cycle. Since the complement of each simplex in L^r is a cycle it follows, conversely, that $\Sigma \bar{c}_j \bar{D}_j^{r-p}$ is a cycle if $\Sigma \bar{c}_j \bar{A}_j^p$ is a co-cycle. If $\Sigma c_i A_i^p$ is a co-cycle, then $\Sigma c_i D_i^{r-p+1}$ is bounded by the dual of $\Sigma b_i A_i^p$ in $\dot{E}_1^{r+1}$, where $b_i = c_i$ if $A_i^p \subset \dot{E}_1^{r+1}$, $b_i = 0$ otherwise. Let $\bar{C}^{r-p}$, given by (3.1), be a cycle, let $Z^p = \Sigma z_i A_i^p$ be an arbitrary cycle in E_1^{r+1}, and let

$$\psi_1(Z^p) = \Sigma \bar{z}_j \bar{A}_j^p = \bar{Z}^p,$$

say, where $\bar{z}_j = \Sigma z_i$, summed for all values of i such that $\psi_1(A_i^p) = \bar{A}_j^p$. Then $\bar{Z}^p$ is a cycle, and $\bar{Z}^p \sim 0$ since $Z^p \sim 0$. It follows from the definition of c_i and $\bar{z}_j$ that

$$\Sigma c_i z_i = \Sigma \bar{c}_j \bar{z}_j,$$

and $\Sigma \bar{c}_j \bar{z}_j = 0$ since $\Sigma \bar{c}_j \bar{A}_j^p$ is a co-cycle. Since the cycle Z^p is arbitrary it follows that $\Sigma c_i A_i^p$ is a co-cycle. Therefore $\psi_1^{-1}(\bar{C}^{r-p})$, given by (3.2), is bounded by the dual of $\Sigma b_i A_i^p$ in $\dot{E}'$, where the coefficients b_i are defined as before. It follows from the analogue of (3.2), for the map ϕu, that the dual of $\Sigma b_i A_i^p$ in $\dot{E}'$ is $u^{-1}\phi^{-1}(\bar{C}^{r-p})$. Therefore

$$u^{-1}\phi^{-1}(\bar{C}^{r-p}) \sim 0 \text{ in } |\psi_1^{-1}(\bar{C}^{r-p})|,$$

and it follows from our previous arguments that $\alpha \neq 0$ if $k-1 \leqslant r-k$ $(r > k)$.

† Since the coefficients are residue classes (mod 2), we do not distinguish between the algebraic complex $\bar{C}^{r-p}$ and the absolute complex $|\bar{C}^{r-p}|$.

We next take $1<r-k<k-1$. The manifold $K_0^{r-1}=u^{-1}\phi^{-1}(K^{r-1})$, where K^{r-1} means the same as in Theorem 4, is a "transversal" of the join $S^{k-1}S_0^{r-k}$, and may therefore be represented as the topological product $S^{k-1}\times S_0^{r-k}$. Since $k-1>1$ and $r-k>1$ it is simply connected. Therefore it is a copy of the universal covering space of K^{r-1}, whence

$$\pi_{r-k}(K^{r-1})=\pi_{r-k}(S^{k-1}\times S_0^{r-k})$$

and is cyclic infinite, since† $r-k<k-1$. Let α_1 be the generator‡ of $\pi_{r-k}(K^{r-1})$, with $\phi u(x_0\times y_0)$ as base point, which is represented by the map $\phi u(x_0\times S_0^{r-k})\subset K^{r-1}$, for any points $x_0\in S^{k-1}$, $y_0\in S_0^{r-k}$. The covering $\phi u(S^{k-1}\times S_0^{r-k})=K^{r-1}$ is 2-sheeted and $u(x,\ y)=u(x^*,\ y^*)$, where $x,\ x^*\subset S^{k-1}$ and $y,\ y^*\subset S_0^{r-k}$ are diametrically opposite pairs of points. Let $t(S^{k-1}\times S_0^{r-k})=S^{k-1}\times S_0^{r-k}$ be the transformation which is given by $t(x,\ y)=(x^*,\ y^*)$. If $r-k$ is even we have $t(x_0\times S_0^{r-k})=-x_0{}^*\times S_0^{r-k}$, taking orientation into account, whence the map $\phi ut(x_0\times S_0^{r-k})$ represents the element $-\alpha_1$. If $r-k$ is odd, then $t(x_0\times S_0^{r-k})=x_0{}^*\times S_0^{r-k}$ and $\phi ut(x_0\times S_0^{r-k})$ represents α_1. In either case, let Γ_2 be the residue group $\pi_{r-k}(K^{r-1})-2\pi_{r-k}(K^{r-1})$, of order 2, where $2\pi_{r-k}(K^{r-1})$ is the sub-group which is generated by $2\alpha_1$. The automorphism of $\pi_{r-k}(K^{r-1})$ in which $\alpha_1\to\epsilon\alpha_1$, for a given value of $\epsilon=\pm1$, clearly leaves $2\pi_{r-k}(K^{r-1})$ invariant and therefore determines an automorphism of Γ_2, namely the identity. Thus a unique element $\bar{\alpha}\in\Gamma_2$ is represented by a free map $f(\Sigma^{r-k})\subset K^{r-1}$, where Σ^{r-k} is an oriented, topological $(r-k)$-sphere, and both orientations of Σ^{r-k} give the same element $\bar{\alpha}$. If $\Sigma^{r-k}\subset K_0^{r-1}$ links S^{k-1} with coefficient 1 it is clear that the map $\phi u(\Sigma^{r-k})\subset K^{r-1}$ represents the non-zero element§ in Γ_2. Finally, notice that a map $f(\Sigma^l)\subset K^{r-1}$ is inessential if $l<r-k$.

Let $\psi_1(E_1^{r+1})\subset P_{k-1}^r$ mean the same as before, let $k+t$ be the maximum dimensionality of the simplexes in $\psi_1^{-1}(\mathfrak{E}^0)$ and let A^{k+t} be any one with this dimensionality. If $t\geqslant0$, then A^{k+t} is internal to E_1^{r+1}, since

† Cf. Note 2 in § 8 below.

‡ If $r-k=1$, then $\alpha_1=2\alpha_0$, for a certain $\alpha_0\in\pi_1(K^{r-1})$.

§ Treated as a cycle, $u(x_0\times S^{r-k})\sim0$ (mod 2) in K_1^{r-1}. For $u(x_0\times S_0^{r-k})$ is the intersection of K_1^{r-1} and a projective $(r-k+1)$-space of the form pP_0^{r-k}, where $p\in P^{k-1}$. Since $r-k+1\leqslant k-1$, there is a projective $(r-k+1)$-space P_0^{r-k-1} (*e.g.* a space $P_0^{r-k+1}\subset P^{k-1}$), which does not meet K_1^{r-1}. Since P_0^{r-k+1} is a basis for the $(r-k+1)$-cycles (mod 2) in P^r there is a chain $C^{r-k+2}\subset P^r$, whose boundary is $P_0^{r-k+1}+pP_0^{r-k}$, and $u(x_0\times S_0^{r-k})$ bounds the intersection $K^{r-1}.C^{r-k+2}$. On the other hand $u(x_0\times S^{r-k})$ does not bound in K^{r-1} if $r-k\geqslant k-1$. For let $r-k\geqslant k-1$ and let $f(S^{k-1})\subset S_0^{r-k}$ be a map such that $f(x^*)=f(x)^*$. Then $\phi u\{x,\ f(x)\}=\phi u\{x^*,\ f(x^*)\}$, whence the locus $\phi u\{x,f(x)\}$ is a topological projective $(k-1)$-space $\Pi_0^{k-1}\subset K^{r-1}$. It may be verified that Π_0^{k-1} intersects $\phi u(x_0\times S^{r-k})$ with coefficient unity.

$\psi_1^{-1}(\mathfrak{E}^0) \cap \dot{E}_1^{r+1} = S^{k-1}$, and E_1^{r+1} is of the form

$$A^{k+t}\Sigma^{r-k+t} + U^{r+1},$$

where Σ^{r-k+t}, the complement of A^{k+t}, is an $(r-k-t)$-sphere. Since the map ψ_1 is simplicial we have $\psi_1(\Sigma^{r-k+t}) \subset K^{r-1}$. If $t \geqslant 1$ we have

$$r-k-t < r-k,$$

whence the map $\psi_1(\Sigma^{r-k-t})$ is inessential in K^{r-1}. That is to say $\psi_1(\Sigma^{r-k-t})$ can be extended to a map $\psi'(a\Sigma^{r-k+t}) \subset K^{r-1}$, where a is a new vertex inside A^{k+t}, and we may assume ψ' to be simplicial with respect to some subdivision $\sigma(a\Sigma^{r-k+t})$ of $a\Sigma^{r-k+t}$. We extend ψ' throughout the simplicial neighbourhood of a by defining

$$\psi'(\dot{A}^{k+t}B) = \mathfrak{E}^0\psi'(B),$$

for each simplex $B \subset \sigma(a\Sigma^{r-k+t})$. We then replace ψ_1 by the map which coincides with ψ' in the neighbourhood of a, and with ψ_1 in the rest of E_1^{r+1}. Thus, by a subdivision and modification of the map ψ_1, we have reduced the number of simplexes in $\psi_1^{-1}(\mathfrak{E}^0)$. It follows from induction on this number that $\psi_1(E_1^{r+1})$ may be replaced by a (simplicial) map $\psi_2(E_2^{r+1})$, such that $\psi_2^{-1}(\mathfrak{E}^0)$ is at most k-dimensional, where E_2^{r+1} is an internal subdivision of E_1^{r+1}, and $\psi_2 = \psi_1$ in $\dot{E}_2^{r+1}$. Let $A_1^k, \ldots, A_s^k$ be the k-simplexes in $\psi_2^{-1}(\mathfrak{E}^0)$, let Σ_i^{r-k} be the complement of A_i^k $(i = 1, \ldots, s)$, and let $\bar{a}_i \in \Gamma_2$ be the element which is represented by the map

$$\psi_2(\Sigma_i^{r-k}) \subset K^{r-1}.$$

I say that no $(k-1)$-simplex, A^{k-1}, which is internal to E_2^{r+1}, is a boundary simplex of the chain

$$C^k = \sum_{i=1} \bar{a}_i A_i^k.$$

For let $A_{i_1}^k, \ldots, A_{i_p}^k$ be those of the k-simplexes in $\psi_2^{-1}(\mathfrak{E}^0)$ which are incident with A^{k-1}, and let Σ^{r-k+1} be the complement of A^{k-1} in E_2^{r+1}. It is trivial that $A^{k-1} \notin \dot{C}^k$ if $p = 0$. If $p > 0$, let Σ_j^{r-k} $(j = i_1, \ldots, i_p)$ be oriented so that, taking account of orientation, $\Sigma_{i_1}^{r-k} + \ldots + \Sigma_{i_p}^{r-k} \sim 0$ in the part N^{r-k+1}, of Σ^{r-k+1}, which lies in $\psi_2^{-1}(K^{r-1})$. Let a given point $b \in N^{r-k+1}$ be joined by a segment, $s_j \subset N^{r-k+1}$, to an arbitrary point in Σ_j^{r-k} $(j = i_1, \ldots, i_p)$ and let $\alpha_j \in \pi_{r-k}(K^{r-1})$ be the element which is represented by the map† $\psi_2(s_j + \Sigma_j^{r-k})$, when $\psi_2(b)$ is taken as the base point for $\pi_{r-k}(K^{r-1})$. Then

† Cf. J. H. C. Whitehead, *Proc. London Math. Soc.* (2), 45 (1939), 280.

$a_{i_1}+\ldots+a_{i_p}=0$. Therefore $\bar{a}_{i_1}+\ldots+\bar{a}_{i_p}=0$, whence† $A^{k-1}\not\subset \dot{C}^k$. Now let $A^{k-1}\subset S^{k-1}$ and let Σ^{r-k} be its complement in $\dot{E}_2^{r+1}$. Since Σ^{r-k} links S^{k-1} with coefficient 1, the map $\psi_2(\Sigma^{r-k})\subset K^{r-1}$ represents the non-zero element in Γ_2. Therefore it follows from a similar argument to the one just given that A^{k-1} appears in $\dot{C}^k$ with the coefficient 1. Therefore $\dot{C}^k$ is the sub-complex of $\dot{E}_2^{r+1}$ which covers S^{k-1}. We now arrive at a contradiction in the same way as when $r\geqslant 2k-1$, replacing the k-cycle $\psi_1^{-1}(\bar{C}_1^{k-1})$ by C^k. Therefore $\alpha\neq 0$ if $r>k+1$ and $1<r-k<k-1$.

It remains to prove that $\alpha\neq 0$ when $r=k+1$ and is even ($r>4$). Let $w(V_{k+2,3})=V_{k+2,2}$ be the map which is given by $w(v_1, v_2, v_3)=(v_2, v_3)$. Then

$$w^{-1}(a_2, a_3)=V_{k,1}=S^{k-1}$$

and, as in the proof of Theorem 14, in § 7 below, it follows from Lemma 4, in § 7 below, that a map $f_0(S^{k+1})\subset V_{k+2,3}$, such that $wf_0(S^{k+1})\subset V_{k+2,2}$ is inessential, is homotopic to a map $f_1(S^{k+1})\subset w^{-1}(a_2, a_3)=S^{k-1}$. Since‡ $(k-1^{k+1})=0$ if $k>3$, it follows that, if $wf_0(S^{k+1})\subset V_{k+2,2}$ is inessential, so is $f_0(S^{k+1})\subset V_{k+2,3}$. If $f_0=\phi u$ it is clear that wf_0 is the corresponding map $\phi u(S^{k+1})\subset P_{k-1}^{k+1}\subset V_{k+2,2}$. If the latter were inessential, then $\phi u(S^{k+1})\subset P_{k-2}^{k+1}\subset V_{k+2,3}$ would be inessential in $V_{k+2,3}$. But $k+2<2k-1$, since $k>3$, whence $\phi u(S^{k+1})\subset P_{k-2}^{k+1}$ would be inessential in P_{k-2}^{k+1}, as in Theorem 3 contradicting what we have proved for

$$r-k'=(k+1)-(k-1)=2.$$

Therefore $\alpha\neq 0$ and the proof is complete.

Except when $r=4$, $k=2$, the argument used in proving Theorem 5 for $r-k\geqslant k-1$ may be simplified as follows. In order to display the parameter k let us write $\phi(P^r)\subset P_{k-1}^r\subset V_{r+1,m}$ $(m=r-k+1)$ as

$$\phi_k(P^r)\subset P_{k-1}^r,$$

and let $w(V_{r+1,m})=V_{r+1,l}$ be the map which is given by

$$w(v_1, \ldots, v_m)=(v_{m-l+1}, \ldots, v_m).$$

Clearly $w(P_{k-1}^r)=P_{h-1}^r$ where $h=r-l+1$, and $w\phi_k=\phi_h$. Therefore, if $\phi_k u(S^r)\subset P_{k-1}^r$ were inessential, then $\phi_h u(S^r)\subset P_{h-1}^r$ would also be

† Cf. S. Eilenberg, *Fund. Math.*, 31 (1938), 170–200.

‡ See L. Pontrjagin, *Report of the Oslo Congress* (1936), and H. Freudenthal, *Comp. Math.*, 5 (1937), 299–314 (footnote on p. 301); *Proc. Akad. Amsterdam*, 42 (1939), 139–140. I have not seen a proof of this theorem. Freudenthal's paper in *Comp. Math.* will be referred to as F.S. As in F.S., we denote $\pi_s(S^d)$ by (d^s).

inessential for $k < h \leqslant r$. When we have proved that $\phi_h u(S^r)$ is essential for† $r-h \leqslant h-1$, it follows that $\phi_k u(S^r)$ is essential for $k = 1, \ldots, h$.

4. *On symmetric maps.* Let X be any arcwise connected topological space, which is r-simple for a given value of r. Let $\Gamma_2 = \pi_r(X) - 2\pi_r(X)$, where $2\pi_r(X)$ is the sub-group consisting of all elements of the form 2β $[\beta \in \pi_r(X)]$. Let $\psi_1(P^r), \psi_2(P^r) \subset X$ be given maps and let $f_\rho = \psi_\rho u$ $(\rho = 1, 2)$, where $u(S^r) = P^r$ means the same as before. Let $\alpha_\rho \in \pi_r(X)$ be the element which is represented by the map $f_\rho(S^r) \subset X$, and let $\bar{\alpha}_\rho$ be the corresponding element of Γ_2. Let $P^{r-1} \subset P^r$ mean the same as before.

THEOREM 6. *If the maps $\psi_1(P^{r-1})$, $\psi_2(P^{r-1})$ belong to the same homotopy class, then $\bar{\alpha}_1 = \bar{\alpha}_2$.*

Let $\psi_2(P^{r-1})$ be deformable into $\psi_1(P^{r-1})$. A deformation of $\psi_2(P^{r-1})$ into $\psi_1(P^{r-1})$ may be extended to a deformation of the whole map‡ $\psi_2(P^r)$, and we therefore assume that $\psi_1 = \psi_2$ in P^{r-1}. Then $f_1 = f_2$ in

$$S^{r-1} = u^{-1}(P^{r-1}).$$

Let $S^r = S_1{}^r \subset R^{r+1}$ be given by $x_0{}^2 + \ldots + x_r{}^2 = 1$, let $S_2{}^r \subset R^{r+1}$ be the quadric $x_0{}^2 + \ldots + x_{r-1}^2 + 2x_r{}^2 = 1$ and let S^{r-1} be the $(r-1)$-sphere, in the hyperplane $x_r = 0$, in which $S_1{}^r$ meets $S_2{}^r$. Let $u_1 = u$ and let $u_2(S_2{}^r) = P^r$ be the resultant of the radial projection of $S_2{}^r$ on $S_1{}^r$, from $(0, \ldots, 0)$, followed by u_1. Then $\bar{\alpha}_\rho$ is equally well represented by either of the maps $g_\rho(S_\rho{}^r) \subset X$ $(\rho = 1, 2)$, where $g_\rho = \psi_\rho u_\rho$. Let $Q^r = S_1{}^r + S_2{}^r$ and let $g(Q^r) \subset X$ be the map which is given by $g = g_\rho$ in $S_\rho{}^r$, this being single-valued since $g_1 = g_2$ in S^{r-1}. Let $E^r_{\rho 1}$ and $E^r_{\rho 2}$ be the upper and lower hemispheres on $S_\rho{}^r$, having S^{r-1} as their common boundary. Let $\bar{\beta}_\sigma \in \Gamma_2$ be the element which is represented by the map

$$g(E^r_{1\sigma} + E^r_{2\sigma}) \subset X,$$

in which orientations are ignored. If $r = 1$ it is clear that $\alpha_\rho = 2\alpha_\rho'$, where α_ρ' is represented by the map $\psi_\rho(P^1)$ and $\pi_1(X)$ is written additively. Therefore $\bar{\alpha}_1 = \bar{\alpha}_2 = 0$. If $r > 1$, it follows from an argument similar to one used in proving Theorem 5, that

$$\bar{\alpha}_1 + \bar{\alpha}_2 = \bar{\beta}_1 + \bar{\beta}_2.$$

† The argument given in the theorem for the case $r-k < k-1$ is valid for $r-k = k-1$ if the residue group $\Gamma_2 = \pi_{r-k}(K^{r-1}) - 2\pi_{r-k}(K^{r-1})$ is replaced by $\Gamma_2 = \pi_{r-k}(K^{r-1}) - (2\alpha_1, \beta)$, where α_1, β are the elements which correspond to the maps $\phi u(x_0 \times S_0{}^{r-k})$, $\phi u(S^{k-1} \times y_0) \subset K^{r-1}$ and $(2\alpha_1, \beta)$ is the sub-group generated by $2\alpha_1$ and β.

‡ P. Alexandroff and H. Hopf, *Topologie* (Berlin, 1935), 501.

Since $g_\rho = \psi_\rho u_\rho$ we have $g_\rho(x_0, \ldots, x_r) = g_\rho(-x_0, \ldots, -x_r)$, whence

$$g(x_0, \ldots, x_r) = g(-x_0, \ldots, -x_r).$$

Therefore $$g(E^r_{11}+E^r_{21}) = gt(E^r_{12}+E^r_{22}),$$

where t is the reflexion in the origin, whence $\bar\beta_1 = \bar\beta_2$ and $\bar\alpha_1+\bar\alpha_2 = 0$. Therefore $\bar\alpha_1 = \bar\alpha_2$ and the theorem is established.

Let us describe a map $f(S^r) \subset X$ as *symmetric* if $f(x) = f(x^*)$ for each $x \in S^r$. Clearly f is symmetric if, and only if, it is of the form ψu, where ψ maps $P^r = u(S^r)$ in X. Since $\bar\alpha_\rho = \alpha_\rho$ if $2\pi_r(X) = 0$ we have the corollary:

COROLLARY. *If $2\pi_r(X) = 0$ then all the symmetric maps $f(S^r) \subset X$, for which the maps $\psi(P^{r-1}) \subset X$ $(f = \psi u)$ belong to a given homotopy class, are themselves in the same homotopy class.*

THEOREM 7. *No symmetric map $f(S^r) \subset S^{r-1}$ is essential unless $r \equiv 3$ (mod 4). If $r \equiv 3$ (mod 4), and if $\psi(P^{r-1}) \subset S^{r-1}$ is essential $(f = \psi u)$, then $f(S^r) \subset S^{r-1}$ is essential. If $r = 3$ and $\psi(P^2)$ is essential, then $f(S^3) \subset S^2$ corresponds to the non-zero element in $\Gamma_2 = (2^3)-2(2^3)$.*

First let r be even, and since any map $f(S^2) \subset S^1$ is inessential, let $r > 3$. Since $r-1$ is odd, P^{r-1} is orientable. Let P^{r-1} and S^{r-1} be oriented and let d be the degree with which $\psi(P^{r-1})$ covers S^{r-1}. Then $f(S^{r-1})$ covers S^{r-1} with degree $2d$. But $f(S^{r-1})$ is part of the map $f(E^r) \subset S^{r-1}$, where E^r is one of the hemispheres into which S^r is separated by S^{r-1}. Therefore $f(S^{r-1}) \subset S^{r-1}$ is inessential, whence $2d = 0$. Therefore $\psi(P^{r-1})$ is homotopic to a constant map† $\psi(P^{r-1}) = p_0 \in S^{r-1}$. Since $r > 3$ we have $2(r-1^r) = 0$, and it follows from the corollary to Theorem 6 that $f(S^r)$ is in the same homotopy class as the constant map $f_1(S^r) = p_0$.

Now let $r = 2l+1$, where $l > 0$, the theorem being trivial if $r = 1$. Then P^{r-1} is non-orientable and there are exactly two homotopy classes of maps $\psi(P^{r-1}) \subset S^{r-1}$, the essential ones being those with the degree 1 (mod 2). If $\psi(P^{r-1})$ is inessential it follows from Theorem 6 that $\bar\alpha = 0$, where $\bar\alpha \in \Gamma_2 = (r-1^r)-2(r-1^r)$ is the element which corresponds to $f(S^r)$, and hence that $f(S^r)$ is inessential if $r > 3$, since $(r-1^r)$ is then of order 2. Since $(r-1^r)$ is cyclic if $r \geqslant 3$, the group Γ_2 is of order 2, and the theorem will follow from Theorem 6 when we have proved that some particular symmetric map $f(S^r) \subset S^{r-1}$ represents the non-zero element $\bar\alpha \in \Gamma_2$ if l is odd, and that some particular symmetric map $f(S^r)$, for which $\psi(P^{r-1})$

† Alexandroff and Hopf (*loc. cit.*), 509, Theorem 1.

is essential, is inessential if l is even. Let $r = 3$, and let us introduce complex coordinates

$$\xi = x_0 + ix_1, \quad \eta = x_2 + ix_3$$

for R^4. Let S^3 and S^2 be given by $\xi\bar{\xi} + \eta\bar{\eta} = 1$ and $\xi\bar{\xi} + {x_2}^2 = 1$ respectively, and $f(S^3) \subset S^2$ by

$$(4.1) \qquad {x_0}' + i{x_1}' = -2\xi\bar{\eta}, \quad {x_2}' = 1 - 2\eta\bar{\eta} \quad (\xi\bar{\xi} + \eta\bar{\eta} = 1).$$

Then $f^{-1}(q)$, for any point $q \in S^2$, is clearly a great circle of the form

$$\xi = \xi_0 e^{i\theta}, \quad \eta = \eta_0 e^{i\theta},$$

and two great circles in S^3 link each other with coefficient ± 1. Therefore the Hopf invariant† of f is ± 1, whence α generates (2^3) and $\bar{\alpha} \neq 0$, where $\alpha \in (2^3)$ and $\bar{\alpha} \in \Gamma_2$ are the elements which correspond to f.

Now let $r > 3$ and for the moment let r be even or odd. Let

$$\xi = x_{r-3} + ix_{r-2}, \quad \eta = x_{r-1} + ix_r,$$

let S^{r-1}, $S_0^{r-2} \subset S^r$ be the spheres given by $x_r = 0$ and $x_0 = x_r = 0$, and let $g(S^{r-1}) \subset S_0^{r-2}$ be an essential map. According to F.S., Theorem 1, a map $f_0(S^r) \subset S^{r-1}$ is essential if $f_0(E_\rho^r) \subset E_\rho^{r-1}$ $(\rho = 1, 2)$ and $f_0 = g$ in S_0^{r-2}, where E_1^r, E_2^{r-1} are the hemispheres of S^r and S^{r-1} in which $x_0 \geqslant 0$, and‡ E_2^r, E_1^{r-1} the hemispheres in which $x_0 \leqslant 0$. It follows from the preceding paragraph and induction on r that the map given by

$$(4.2) \qquad {x_p}' = -2x_p|\eta|, \quad \xi' = -2\xi\bar{\eta}, \quad x'_{r-1} = 1 - 2\eta\bar{\eta} \quad (p = 0, \ldots, r-4)$$

is essential.

Now let $r = 2l+1$ $(l > 1)$, let

$$\xi_j = x_{2j} + ix_{2j+1}, \quad \eta = x_{2l} + ix_{2l+1} \quad (j = 0, \ldots, l-1)$$

and let $f(S^r) \subset S^{r-1}$ be the (symmetric) map, which is given by

$$(4.3) \qquad {\xi_j}' = -2\xi_j\bar{\eta}, \quad x'_{2l} = 1 - 2\eta\bar{\eta}.$$

Taking $x_r = 0$, that is to say taking η to be real, we see that the map $\psi(P^{r-1}) = S^{r-1}$ coincides with $\phi(P^{r-1}) = S^{r-1}$. The latter is of degree 1 (mod 2) and is therefore essential. Therefore Theorem 7 will follow from

† H. Hopf, *Math. Annalen*, 104 (1931), 637–665. This is the example given by Hopf of a map with the invariant ± 1.

‡ This is slightly awkward, but we want to bring out the analogy between (4.3), below, and the expression for x, a, when $a = (0, \ldots, 0, 1)$.

the corollary to Theorem 6 when we have proved that f is essential if l is odd and inessential if l is even.

Let $$T^l = S_0^1 \times \ldots \times S_{l-1}^1,$$

where S_j^1 is an oriented circle, which is described positively as an angular variable θ_j increases from 0 to 2π. Let G_r be the group of rotations in $R^r \subset R^{r+1}$, given by $x_r = 0$, about the centre of S^{r-1}, and let

$$h\{\pi_1(T^l)\} \subset G_r$$

be the homomorphism which is induced by the map in which

$$(\theta_0, \ldots, \theta_{l-1}) \in T^l$$

corresponds to the rotation

$$\xi_j' = e^{-i\theta_j}\xi_j, \quad x_{2l}' = x_{2l} \quad (j = 0, \ldots, l-1).$$

It follows from Theorem 2 that $h(g_0) = \ldots = h(g_{l-1}) = \beta$, where $g_j \in \pi_1(T^l)$ is the element corresponding to S_j^1, and β is the non-zero element in $\pi_1(G_r)$, written additively. Therefore

$$\begin{aligned} h(g_0 + \ldots + g_{l-1}) &= l\beta \\ &= \beta \quad \text{if } l \text{ is odd} \\ &= 0 \quad \text{if } l \text{ is even.} \end{aligned}$$

Let χ be an angular coordinate for a circle S^1, and let $\rho(\chi) \in G_r$ be the rotation

$$\xi_j' = e^{-i\chi}\xi_j, \quad x_{2l}' = x_{2l}.$$

Then the map $\chi \to \rho(\chi)$ is homotopic in G_r to the map $\chi \to \rho_1(\chi)$, where $\rho_1(\chi)$ is given by

$$\xi_\lambda' = \xi_\lambda, \quad \xi_{l-1}' = e^{-i\chi}\xi_{l-1}, \quad x_{2l}' = x_{2l} \quad (\lambda = 0, \ldots, l-2)$$

if l is odd, and is the identity in G_r if l is even.

Let $\eta = |\eta| e^{i\chi}$ and let $f_1(S^r) \subset S^{r-1}$ be the map which is given by

$$\xi_j' = -2\xi_j|\eta|, \quad x_{2l}' = 1 - 2\eta\bar{\eta}. \tag{4.4}$$

Then the map $f(S^r) \subset S^{r-1}$, which is given by (4.3), may be expressed in the form

$$f = \rho(\chi) f_1.$$

Therefore, if l is odd, it is homotopic to the map f_0, which is given by (4.2), and is essential. If l is even it is homotopic to the map which is given by (4.4). But

$$f_1(x_0, \ldots, x_{r-1}, x_r) = f_1(x_0, \ldots, x_{r-1}, -x_r),$$

since $|\eta| = |\bar{\eta}|$, and it follows from a previous argument that f_1 is inessential. Therefore f is essential or inessential according as l is odd or even, and the proof is complete.

5. We now prove a theorem† concerning the relation between $\pi_r(X)$ and $\pi_r(X+\mathscr{E}^t)$, where $t > 1$ and $\mathscr{E}^t$ is a cell which is attached to an arcwise connected, topological space X. More precisely, let E^t be a t-element, which does not meet X, let $\mathscr{E}^t$ be the interior of E^t, and let $f(\dot{E}^t) \subset X$ be a given map. Then $X^* = X + \mathscr{E}^t$ is the space obtained by the identification of each point $p \in \dot{E}^t$ with $f(p) \in X$. We extend $f(\dot{E}^t) \subset X$ to a map $f(E^t) \subset X^*$ by writing $f(p) = p$ for each point $p \in \mathscr{E}^t$. Let

$$h_s(t-1^s) \subset \pi_s(X)$$

be the homomorphism which is determined by the map $f(\dot{E}^t) \subset X$ and let $\phi_s\{\pi_s(X)\} \subset \pi_s(X^*)$ be the homomorphism in which corresponding elements are represented by the same map. Let S^{r-1}, S^r mean the same as before and let E_1^r, E_2^r be the two r-elements into which S^r is separated by S^{r-1}. Let $g_0(S^{r-1}) \subset \dot{E}^t$ be a map such that $fg_0(S^{r-1}) \subset X$ is inessential, and which can therefore be extended to a map $g_1(E_1^r) \subset X$ ($g_1 = fg_0$ in S^{r-1}). Clearly g_0 can be extended to a map $g_2(E_2^r) \subset E^t$, and we define $g(S^r) \subset X^*$ as the map which is given by

$$(5.1) \qquad \begin{aligned} g &= g_1 \quad \text{in} \quad E_1^r \\ &= fg_2 \quad \text{in} \quad E_2^r. \end{aligned}$$

Let $\alpha \in (t-1^{r-1})$ be the element which is represented by the map g_0, and let $\psi_r(\alpha) \in \pi_r(X^*)$ be the element which is represented by $g(S^r) \subset X^*$, with a suitable choice of base point. In general, $\psi_r(\alpha)$ depends on the choice of the deformation cell $g_1(E_1^r) \subset X$. But it follows from an argument similar to arguments used in the first part of Theorem 5 and in Theorem 6, that any two values of $\psi_r(\alpha)$, which correspond to the same element $\alpha \in (t-1^{r-1})$, differ by an element in $\phi_r\{\pi_r(X)\}$. Therefore the transformation $\alpha \to \psi_r(\alpha)$ is single-valued if $\psi_r(\alpha)$ is interpreted as an element in the residue group $\pi_r(X^*) - \phi_r\{\pi_r(X)\}$. The transformation $\alpha \to \psi_r(\alpha)$ is defined for any $\alpha \in h_{r-1}^{-1}(0)$ and it may be verified that‡ it is a homomorphism of $h_{r-1}^{-1}(0)$ in $\pi_r(X^*) - \phi_r\{\pi_r(X)\}$.

† This theorem overlaps with the combined result of three theorems which I have proved elsewhere (*Proc. London Math. Soc., loc. cit.*, Theorems 18, p. 281, and 19, p. 285, and *Annals of Math.*, 41 (1940), Theorem 3). Each is, in some respects, more general than the other.

‡ The argument is essentially the same as that which shows that $\mathfrak{G}(d-1^{e-1}) \subset (d^e)$ is a homomorphism (F.S.).

THEOREM 8. *If* $\pi_s(X) = 0$ *when* $1 \leqslant s \leqslant r-t+1$, *and if* $r < 2t-1$, *then*

$$(a) \qquad \psi_r\{h_{r-1}^{-1}(0)\} = \pi_r(X^*) - \phi_r\{\pi_r(X)\}.$$

If, in addition, $\pi_{r-t+2}(X) = 0$ *and* $r < 2t-2$, *then*

(b) $\psi_r^{-1}(0) = 0$, *whence* ψ_r *is an isomorphism of* $h_{r-1}^{-1}(0)$ *on the residue group* $\pi_r(X^*) - \phi_r\{\pi_r(X)\}$,

(c) $\phi_r^{-1}(0) = h_r(t-1^r)$.

If $r < t-1$ the theorem is trivial. If $r = t-1$ we have

$$h_{r-1}^{-1}(0) \subset (t-1^{t-2}) = 0$$

and the rest of the theorem has been proved elsewhere†. Therefore we assume that‡ $r \geqslant t$. We first prove (a), by showing that a given element $\alpha \in \pi_r(X^*)$ is represented by a map which also represents an element in $\psi_r\{h_{r-1}^{-1}(0)\}$, and hence that ψ_r is a map of $h_{r-1}^{-1}(0)$ on $\pi_r(X^*) - \phi_r\{\pi_r(X)\}$. Let $g_0(\Sigma^r) \subset X^*$ be a map which represents α, where Σ^r is a rectilinear, simplicial triangulation of the r-sphere, which consists of R^r, together with an ideal point§ (∞). Assume that g_0 is homotopic to a map $g_1(\Sigma^r) \subset X^*$, such that $g_1^{-1}(p)$ is a single point, where $p \in \mathcal{E}^t$. Then, after a further deformation if necessary, we may assume that $g_1(S_0^{r-1}) \subset S_0^{t-1}$, where $S_0^{r-1} \subset R^r$ and $S_0^{t-1} \subset \mathcal{E}^t$ are spheres surrounding $g_1^{-1}(p)$ and p; also that $g_1(\Sigma^r - \mathcal{E}_0^r) \subset X^* - \mathcal{E}_0^t$, where $\mathcal{E}_0^r$ and $\mathcal{E}_0^t$ are the cells in R^r and $\mathcal{E}^t$ which are bounded by S_0^{r-1} and S_0^{t-1}. Moreover, we may take S_0^{r-1} to be the equatorial $(r-1)$-sphere S^{r-1}, in terms of which ψ_r is defined. Let $\rho_\lambda(E^t) = E^t$ $(1 \leqslant \lambda \leqslant 2)$ be a deformation of the identical map $\rho_1(E^t) = E^t$, such that $\rho_\lambda(x) = x$ if $x \in \dot{E}^t$, $\rho_2(S_0^{t-1}) = \dot{E}^t$, and let ρ_λ be extended throughout X^* by taking $\rho_\lambda(x) = x$ if $x \in X$. Then $g_\lambda = \rho_\lambda g_1$ is a deformation of the map g_1 into a map g_2, which represents an element in $\psi_r\{h_{r-1}^{-1}(0)\}$.

So we have to show that $g_0(\Sigma^r) \subset X^*$ is homotopic to a map $g_1(\Sigma^r) \subset X^*$, such that $g_1^{-1}(p)$ is a single point. It is easily proved that we lose no generality in assuming $g_0^{-1}(A^t)$ to be a sub-complex of Σ^r and that the map $g_0\{g_0^{-1}(A^t)\} = A^t$ is simplicial and linear in each simplex of $g_0^{-1}(A^t)$, where A^t is a closed simplex in $\mathcal{E}^t$. Then $g_0^{-1}(p)$ is covered by a sub-complex of the derived complex of Σ^r, assuming that the new vertices in

† J. H. C. Whitehead, *Proc. London Math. Soc., loc. cit.*, Theorem 18.

‡ Since $\pi_1(X) = 0$, whence $\pi_1(X^*) = 0$, since $t > 1$, we need not refer $\pi_r(X)$, $\pi_r(X^*)$, ... to a base point.

§ We take (∞) to be an inner point of an r-simplex, the remaining simplexes in Σ^r being finite.

$g_0^{-1}(A^t)$ are placed in $g_0^{-1}(p)$, where p is any inner point of A^t. Let a be an arbitrary point in $R^r - g_0^{-1}(p)$, let T^{r-t} be the polyhedron in which the cone swept out by the (linear) segment $ax \subset R^r$ $[x \in g_0^{-1}(p)]$ cuts $\{g_0^{-1}(\dot{A}^t)\}$, and let C^{r-t+1} be the cone swept out by the segment ay $(y \in T^{r-t})$. Since $r < 2t-1$, we have $r-t < t-1$. Therefore, if B^q is a q-simplex and $q \leqslant r-t+1$, any map $\theta(\dot{B}^q) \subset \dot{A}^t$ is inessential in $\dot{A}^t$. Therefore a map $\theta(B^q) \subset X^* - \mathcal{A}^t$ $(\mathcal{A}^t = A^t - \dot{A}^t)$, with $\theta(\dot{B}^q) \subset \dot{A}^t$, may be extended to a map of the form $\theta(B^q + E^q) \subset X^* - \mathcal{A}^t$, with $\theta(E^q) \subset \dot{A}^t$, where E^q is a q-element with the same boundary as B^q. Obviously $\pi_q(X^* - \mathcal{A}^t) = \pi_q(X) = 0$, since $q \leqslant r-t+1$, and it follows that $\theta(B^q)$ can be transformed into a map $\theta_1(B^q) \subset \dot{A}^t$ by a deformation in $X^* - \mathcal{A}^t$, throughout which $\theta(x)$ is held fixed, for each point $x \in \dot{B}^q$. Therefore a familiar argument shows that $g_0(C^{r-t+1})$ can be transformed into a map $g_0'(C^{r-t+1}) \subset \dot{A}^t$, without disturbing the map $g_0(T^{r-t})$. The rest of the argument refers only to Σ^r and a neighbourhood of p in $\mathcal{E}^t$ and is the same as when X^* is a t-sphere†.

We now prove (b). Let $g_0(S^{r-1}) \subset \dot{E}^{t-1}$ be a map which represents an element $\alpha \in \psi_r^{-1}(0)$. Since $\psi_r(\alpha) = 0$ the map $g(S^r) \subset X^*$, which is given by (5.1), can be extended to a map $g(C^{r+1}) \subset X^*$, where $C^{r+1} = S^r \times \langle 0, 1 \rangle$ $(S^r \times 0 = S^r)$, such that $g(S^r \times 1) \subset X$. Assuming that $r-t+1 < t-1$ and $\pi_{r-t+2}(X) = 0$, it follows from an argument which is similar to the one in the preceding paragraph‡, that $g(C^{r+1})$ is homotopic to a map $\bar{g}(C^{r+1}) \subset X^*$, such that $\bar{g}^{-1}(p)$ is a single point in E_2^r, where $p \in \mathcal{E}^t$ and E_2^r means the same as in (5.1). Moreover, no point

$$g(x) \in \bar{\mathcal{E}}^t = \mathcal{E}^t + f(\dot{E}^t),$$

where $x \in E_2^r$, need be taken outside $\bar{\mathcal{E}}^t$ during the deformation $g \to \bar{g}$, and $g(x)$ need not be moved at all if x is in a suitably chosen neighbourhood of $S^{r-1} = \dot{E}_2^r$. Therefore we assume that $g = \bar{g}$ in some neighbourhood of S^{r-1}, and that $\bar{g}(E_2^r) \subset \bar{\mathcal{E}}^t$. It follows from the continuity of $\bar{g}$ that $\bar{g}(U) \subset \mathcal{E}^t$, where U is some neighbourhood, in C^{r+1}, of $\bar{g}^{-1}(p)$. Therefore S^{r-1} bounds an r-element

$$E_3^r \subset \bar{g}^{-1}(\bar{\mathcal{E}}^t) - \bar{g}^{-1}(p) = \bar{g}^{-1}(\bar{\mathcal{E}}^t - p),$$

which, we may assume, coincides with E_2^r near S^{r-1}. Let us also assume, as we obviously may, that $g_2(E_2^r - S^{r-1}) \subset \mathcal{E}^t$, where $g_2(E_2^r) \subset E^t$ means

† See F.S., p. 309. Here, and in proving that $\psi_r^{-1}(0) = 0$, we only need the easier part of Freudenthal's argument, since $r-t+1 < t-1$.

‡ One can base this argument on a representation of $C^{r+1} - (S^r \times 1)$ as a set of the form $E^{r+1} - q \subset R^{r+1}$, where E^{r+1} is a (closed) convex cell, of which q is an inner point.

the same as in (5.1). Then, since $f(x) = x$ if $x \in \mathfrak{E}^t$, whence $g = fg_2 = g_2$ in $E_2{}^r - S^{r-1}$, and since $\bar{g} = g$ near S^{r-1}, it follows that $\bar{g} = g_2$ in $E_2{}^r$ near S^{r-1}. Therefore the map $g'(E_3{}^r) \subset E^t - p$ is continuous, where

$$\begin{aligned} g' &= g_2 \quad \text{in} \quad S^{r-1} \\ &= \bar{g} \quad \text{in} \quad E_3{}^r - S^{r-1}. \end{aligned}$$

Therefore $g_2(S^{r-1})$ bounds the cell $g'(E_3{}^r) \subset E^t - p$, and is therefore inessential in $\dot{E}^t$. But $g_2 = g_0$ in S^{r-1}, whence $\alpha = 0$. Thus part (b) of the theorem is established.

Part (c) also follows from a similar argument. For let $g(S^r) \subset X$ be a map which represents a given element $\alpha \in \phi_r^{-1}(0)$. Since $\alpha \in \phi_r^{-1}(0)$, the map $g(S^r)$ can be extended to a map $g(E^{r+1}) \subset X^*$, where $S^r = \dot{E}^{r+1}$. Since $r-t+1 < t-1$ and $\pi_s(X^*) = 0$ if $s \leqslant r-t+2$, we may assume that $g^{-1}(p)$ is a single point $q \in E^{r+1} - S^r$ $(p \in \mathfrak{E}^t)$, and that $g(S_0{}^r) = S_0^{t-1}$, where $S_0{}^r \subset E^{r+1}$ and $S_0^{t-1} \subset \mathfrak{E}^t$ are spheres with q and p as their centres. Let $\rho(S^r) = S_0{}^r$ be the radial projection of S^r from q on $S_0{}^r$, and $\rho'(S_0^{t-1}) = \dot{E}^t$ the radial projection of S_0^{t-1} from p on $\dot{E}^t$. Then $g(S^r) \subset X$ is obviously homotopic, in $X^* - p$, first to the map

$$g\rho(S^r) = fg\rho(S^r) \subset S_0^{t-1},$$

since $f(x) = x$ if $x \in \mathfrak{E}^t$, and then to

$$f\rho' g\rho(S^r) \subset f(\dot{E}^t).$$

Since the maps f and $f\rho' g\rho$ are homotopic in $X^* - p$ they are homotopic in X. The map $f\rho' g\rho$ represents an element in $h_r(t-1^r)$, whence $\phi_r^{-1}(0) \subset h_r(t-1^r)$. It is obvious that $h_r(t-1^r) \subset \phi_r^{-1}(0)$, and the theorem follows.

Let the map $f(\dot{E}^t) \subset X$ be inessential in X. Then it can be extended to a map $f(E_0{}^t) \subset X$, where $E_0{}^t$ is a hemisphere of a sphere S^t, of which E^t is the other hemisphere. Let $h_r{}^*(t^r) \subset \pi_r(X^*)$ be the homomorphism which is determined by the map $f(S^t) \subset X^*$, where $f(x) = x$ if $x \in \mathfrak{E}^t$. Then the homomorphism $\psi_r(t-1^{r-1}) \subset \pi_r(X^*) - \phi_r\{\pi_r(X)\}$ may be replaced by the homomorphism

$$h_r{}^* \mathfrak{G}(t-1^{r-1}) \subset \pi_r(X^*),$$

where $\mathfrak{G}$ means the same as in F.S. Clearly each element in $h_r{}^* \mathfrak{G}(t-1^{r-1})$ belongs to a residue class $\psi_r(\alpha)$, whence each element in this residue class is expressed uniquely in the form

$$h_r{}^* \mathfrak{G}(\alpha) + \beta,$$

where $\beta \in \phi_r\{\pi_r(X)\}$. Also $h_r(t-1^r) = 0$, whence $\phi_r^{-1}(0) = 0$, subject to the conditions of Theorem 8 (*b*) and (*c*). Therefore we have the corollary:

COROLLARY. *Subject to the conditions of Theorem* 8 (*b*) *and* (*c*), *if the map* $f(\dot{E}^t)$ *is inessential in* X, *then*†

$$\pi_r(X^*) = h_r^* \mathfrak{G}(t-1^{r-1}) + \phi_r\{\pi_r(X)\} \simeq (t^r) + \pi_r(X).$$

Now let X be any arc-wise connected topological space. Let $f(S^t) \subset X$ be a map which represents an element $\alpha \in \pi_t(X)$, with base points $p_0 \in S^t$ and $x_0 = f(P_0) \in X$, let $g(S^r) \subset S^t$ represent a given element $\beta \in (t^r)$, with a base point $q_0 = g^{-1}(p_0) \in S^r$, and let $\gamma \in \pi_r(X)$ be the element which is represented by the map $fg(S^r) \subset X$. Clearly γ depends only on the elements α, β and may be written as $\gamma = \alpha(\beta)$.

THEOREM 9. *The transformation* $\beta \to \alpha(\beta)$ *is a homomorphism of* (t^r) *in* $\pi_r(X)$. *If* $r < 2t-1$, *or if* t *is odd and* $r = 2t-1$, *then the transformation* $\alpha \to \alpha(\beta)$ *is a homomorphism of* $\pi_t(X)$ *in* $\pi_r(X)$.

The first part is trivial, since $\beta \to \alpha(\beta)$ is obviously the homomorphism which is induced by f. The second part is trivial if $r = t = 1$. For then $\alpha(\beta) = d\beta$, where d is the degree of the map $g(S^1) = S^1$. It is also trivial if $t = 1$, $r > 1$, or if $r < t$, since then $\beta = \alpha(\beta) = 0$. We therefore assume that $r \geqslant t > 1$. Clearly $0(\beta) = 0$, and consequently we have to show that

$$(\alpha_1 + \alpha_2)(\beta) = \alpha_1(\beta) + \alpha_2(\beta),$$

for given elements α_1, $\alpha_2 \subset \pi_t(X)$. Let S^t be separated into two hemispheres E_1^t and E_2^t with a common boundary $S^{t-1}(p_0 \in S^{t-1})$, and let α_i be represented by a map $f_i(S^t) \subset X$, such that $f_i(E_j^t) = x_0$ $(j \neq i)$, and let $\alpha_1 + \alpha_2$ be represented by the map f, which coincides with f_i in E_i^t. Let S^r be separated into two hemispheres E_1^r and E_2^r, with a common boundary $(q_0 \in \dot{E}_i^r)$. Since either $r < 2t-1$, or t is odd and $r = 2t-1$, it follows from F.S., Theorem 1, that the map $g(S^r) \subset S^t$ is deformable into a map g_1, such that $g_1(E_i^r) \subset E_i^t$ $(i = 1, 2)$, and we may assume that $g(q_0)$ is held fixed throughout the deformation $g \to g_1$. Since g_1 is homotopic to g the element $(\alpha_1 + \alpha_2)(\beta)$, corresponding to the map fg, is also represented by the map fg_1. Similarly, $\alpha_i(\beta)$ is represented by $f_i g_1$. But $f_i g_1(E_j^r) = f_i(E_j^t) = x_0$ $(j \neq i)$ and $fg_1 = f_i g_1$ in E_i^r. Therefore $(\alpha_1 + \alpha_2)(\beta) = \alpha_1(\beta) + \alpha_2(\beta)$, and the proof is complete.

6. We are now ready to calculate $\pi_{k+1}(V_{k+m,m})$ and $\pi_{k+2}(V_{k+m,m})$. We use Z_g to stand for a cyclic group of order g $(g = 2, \ldots, \infty)$, and write

$$\pi_{k+p}(V_{k+m,m}) = \pi_{km}^p.$$

† The symbol $\simeq$ denotes isomorphism.

THEOREM 10. *The groups π^1_{km} are given by the following table, in which the first column indicates the various values of k, and $l \geqslant 1$.*

	$m=1$	$m=2$	$m \geqslant 3$
$k=1$	0	0	0
2	Z_2	$Z_\infty + Z_\infty$	Z_∞
$2l+1$	Z_2	Z_2	0
$4l$	Z_2	$Z_2 + Z_\infty$	$Z_2 + Z_2$
$4l+2$	Z_2	$Z_2 + Z_\infty$	Z_4

The row $\pi^1_{1m} = 0$ follows from the fact that $V_{m+1,m}$ is homeomorphic to the rotation group G_{m+1}, and $\pi_2(G_{m+1}) = 0$. The column under $m=1$ follows from the fact that $V_{k+1,1} = S^k$. Since S^3 admits a parallelism it follows that $V_{4,2}$ may be represented as the topological product† $S^3 \times S^2$, whence

$$\begin{aligned}\pi^1_{22} &= (3^3) + (2^3) \\ &= Z_\infty + Z_\infty.\end{aligned}$$

Let $k > 1$. Then $k+1 < 2k$, whence

$$\pi^1_{k2} = \pi_{k+1}(P^{k+1}_{k-1}), \quad \pi^1_{km} = \pi^1_{k3} = \pi_{k+1}(P^{k+2}_{k-1}) \quad (m \geqslant 3),$$

according to Theorem 3. We have

$$P^{k+2}_{k-1} = S^k + \mathcal{E}^{k+1} + \mathcal{E}^{k+2},$$

as in § 2, where $\mathcal{E}^{k+p}$ is bounded by the map

$$\phi u(S^{k+p-1}) \subset P^{k+p-1}_{k-1} \quad (p = 1, 2).$$

Also $(k+1)-(k+p)+2 = 3-p$, $k+1 < 2(k+p)-2$ and $\pi_s(P^{k+p-1}_{k-1}) = 0$ for $1 \leqslant s \leqslant 3-p$, provided that $p = 2$ when $k = 2$.

To calculate π^1_{k2}, when $k > 2$, let $r = t = k+1$, $X = P^k_{k-1}$ and $f = \phi u$ in Theorem 8. Since $\phi u(S^k) = P^k_{k-1}$ is of degree two or zero, according as k is odd or even, and since (k^k) is a free cyclic group, it follows that

$$\begin{aligned}(6.1) \qquad h_k^{-1}(0) &= 0 \quad \text{if } k \text{ is odd} \\ &= (k^k) \quad \text{if } k \text{ is even},\end{aligned}$$

† Cf. (6.5) below, and note 2 in § 7 below.

and it follows from Theorem 8 (a), and the corollary, that

$$\pi^1_{k2} = \phi_{k+1}(k^{k+1}) \quad \text{if } k \text{ is odd} \tag{6.2}$$

$$= \phi_{k+1}(k^{k+1}) + \psi_{k+1}(k^k) \quad \text{if } k \text{ is even,}$$

where ϕ_{k+1} means the same as in Theorem 8, and we have written (k^{k+1}) for π^1_{k1}. If k is odd, then $\phi u(S^k) \subset P^k_{k-1}$ is of degree 2, and it follows from Theorem 9 that

$$h_{k+1}(k^{k+1}) \subset 2(k^{k+1}) = 0,$$

whence $\phi^{-1}_{k+1}(0) = 0$, according to Theorem 8 (c). Therefore ϕ_{k+1} is an isomorphism and, since $(k^{k+1}) = Z_2$ if $k > 2$, it follows from (6.2) that

$$\pi^1_{k2} = Z_2 \quad (k = 2l+1).$$

If k is even we have, from (6.2) and the corollary to Theorem 8,

$$\pi^1_{k2} = Z_2 + Z_\infty \quad (k = 2l+2).$$

hus the column under $m = 2$ is complete.

Let $k = 2l+1$, and let $\alpha \in \pi^1_{k2}$ be the element corresponding to the map $\phi u(S^{k+1}) \subset P^{k+1}_{k-1}$. Then it follows from Theorem 8, with

$$r = t-1 = k+1,$$

that

$$\pi^1_{k3} \simeq \pi^1_{k2} - (\alpha),$$

where (α) is the sub-group of π^1_{k2} which is generated by α. It follows from Theorem 5 that $\phi u(S^{k+1}) \subset P^{k+1}_{k-1}$ is essential, whence α represents the non-zero element in π^1_{k2}. Therefore

$$\pi^1_{k3} = \pi^1_{km} = 0 \quad (k = 2l+1,\ m \geqslant 3).$$

Let k be even, let $S^{k+1} = S^{k+1}_1 \subset R^{k+2}$ be given by

$$x_0{}^2 + \ldots + x^2_{k+1} = 1,$$

and $S^k \subset S^{k+1}_1$ by this equation and $x_{k+1} = 0$. Let S^{k+1}_2 be the quadric

$$x_0{}^2 + \ldots + x_k{}^2 + 2x^2_{k+1} = 1,$$

and let $E^{k+1}_{\rho\sigma}$ ($\rho, \sigma = 1, 2$) mean the same as in the proof of Theorem 6 ($x_{k+1} \geqslant 0$ in $E^{k+1}_{\rho 1}$). Since k is even, the map $\phi u(S^k) \subset P^k_{k-1} = S^k$ is inessential, and can therefore be extended to a map $\psi_1(E^{k+1}_{21}) \subset S^k$. Let $\psi_2(E^{k+1}_{22}) \subset S^k$ be the map which is given by $\psi_2 = \psi_1 t$, where

$$t(x_0, \ldots, x_{k+1}) = (-x_0, \ldots, -x_{k+1}),$$

and let $\psi(S_2^{k+1}) \subset S^k$ be given by

$$\psi = \psi_\rho \quad \text{in} \quad E_{2\rho}^{k+1}.$$

Let $E_{2\rho}^{k+1}$ be so oriented that

$$\dot{E}_{2\rho}^{k+1} = -\dot{E}_{1\rho}^{k+1},$$

where $E_{1\rho}^{k+1}$ takes its orientation from a given orientation of S_1^{k+1}. Let $\alpha \in \pi_{k2}^1$ be the element corresponding to the map $\phi u(S^{k+1})$ and this orientation of $S^{k+1} = S_1^{k+1}$. Let $\alpha_1 \in \pi_{k2}^1$ be the element which is represented by the map

$$\psi(E_{11}^{k+1} + E_{21}^{k+1}) \subset P_{k-1}^{k+1},$$

where

$$\psi = \phi u \quad \text{in} \quad E_{11}^{k+1}$$

$$= \psi_1 \quad \text{in} \quad E_{21}^{k+1},$$

and let $\beta \in \pi_{k2}$ be the element which is represented by $\psi(S_2^{k+1}) \subset P_{k-1}^k$. It is clear from Theorem 8 that α_1 generates the summand Z_∞ in π_{k2}^1, or one of them if $k = 2$. It follows, from arguments which are similar to those in Theorem 6 and the first part of Theorem 5, that

$$\alpha = 2\alpha_1 - \beta.$$

Let $k = 2$, and let β_1 be a generating element of the summand Z_∞, in π_{22}^1, which is not the one generated by α_1. Since $\phi(P^k) \subset S^k$ is of degree 1 (mod 2) it follows from Theorem 7 that

$$\beta = (2c+1)\beta_1,$$

for some (integral) value of c. Therefore π_{23}^1 is generated by α_1, β_1, subject to the relation

$$2\alpha_1 - (2c+1)\beta_1 = 0.$$

Therefore

$$\pi_{2m}^1 = \pi_{23}^1 = Z_\infty \quad (m \geqslant 3).$$

Let $k = 4l+2$ $(l \geqslant 1)$. Then it follows from Theorem 7 that the map $\psi(S_2^{k+1}) \subset S^k = P_{k-1}^k$ is essential in P_{k-1}^k. Since $\phi u(S^k) \subset P_{k-1}^k$ is inessential it follows from Theorem 8 (c) that $\psi(S_2^{k+1})$ is also essential in

$$P_{k-1}^{k+1} = P_{k-1}^k + \mathcal{E}^{k+1}.$$

Therefore β is the non-zero element in the summand Z_2 of π_{k2}^1. It follows that π_{k3}' is generated by α_1, β, subject to the relations

$$2\beta = 0, \quad 2\alpha_1 - \beta = 0,$$

whence

$$\pi_{km}^1 = \pi_{k3}^1 = Z_4 \quad (k = 4l+2;\ m \geqslant 3).$$

Finally, if $k = 4l$, then $\psi(S_2^{k+1})$ is inessential, according to Theorem 7, and $\beta = 0$. Therefore π_{k3}^1 is generated by α_1, β_1, where β_1 generates the summand Z_2 of π_{k2}^1, subject to the relations $2\alpha_1 = 2\beta_1 = 0$. Therefore

$$\pi_{km}^1 = \pi_{k3}^1 = Z_2 + Z_2 \quad (k = 4l;\ m \geqslant 3)$$

and the proof is complete.

THEOREM 11. *The groups π_{km}^2 are given by the following table, in which $l \geqslant 1$.*

	$m=1$	$m=2$	$m=3$	$m \geqslant 4$
$k=1$	0	Z_∞	$Z_\infty + Z_\infty$	Z_∞
2	Z_2	Z_2+Z_2	Z_2	0
$2l+1$	0	Z_2	Z_2+Z_∞	Z_∞
$2l+2$	0	Z_2	Z_2	0

The column under $m = 1$ follows from the relation $\pi_{k1}^2 = (k^{k+2})$. We proceed next to the general cases $k = 2l+2$ and $2l+3$. Since $k > 2$ we have $k+2 < 2k$, whence

$$\pi_{k2}^2 = \pi_{k+2}(P_{k-1}^{k+1}), \quad \pi_{k3}^2 = \pi_{k+2}(P_{k-1}^{k+2}), \quad \pi_{km}^2 = \pi_{k+2}(P_{k-1}^{k+3}) \quad (m \geqslant 4),$$

according to Theorem 3. Also $(k+2)-(k+p)+2 = 4-p$ $(1 \leqslant p \leqslant 3)$ and $\pi_s(P_{k-1}^{k+p-1}) = 0$ if $1 \leqslant s \leqslant 4-p \leqslant 3$, since $k > 3$. Therefore we can use Theorem 8 in calculating $\pi_{k+2}(P^{k+p})$. We have

$$P_{k-1}^{k+1} = P_{k-1}^k + \mathcal{E}^{k+1},$$

where $\mathcal{E}^{k+1}$ is bounded by the map $\phi u(S^k) = P_{k-1}^k$, and we take $r = k+2$, $t = k+1$, $X = P_{k-1}^k$ in Theorem 8. Then $(t-1^{r-1}) = (k^{k+1}) = Z_2$. Since $\phi u(S^k)$ is of degree 0 or 2, according as $k = 2l+2$ or $2l+3$, and since $k+1 < 2k-1$, it follows from Theorem 9, with $r = k+1$, $t = k$, that $h_{k+1}(k^{k+1}) = 0$, and hence $h_{k+1}^{-1}(0) = Z_2$. Also

$$\phi_r\{\pi_r(X)\} = \phi_{k+2}(\pi_{k1}^2) = \phi_{k+2}(0) = 0,$$

where ϕ_r means the same as in Theorem 8. Therefore

$$\pi_{k2}^2 = Z_2 \quad (k > 3)$$

by Theorem 8 (a), whether k is even or odd.

Now let $k = 2l+2$. In calculating

$$\pi^2_{k3} = \pi_{k+2}(P^{k+1}_{k-1} + e^{k+2})$$

we apply Theorem 8, with $r = t = k+2$, $X = P^{k+1}_{k-1}$. By Theorem 10,

$$\pi_{k+1}(P^{k+1}_{k-1}) = Z_2 + Z_\infty,$$

and the element represented by $\phi u(S^{k+1}) \subset P^{k+1}_{k-1} = X$ is of the form $2\alpha+\beta$, where α generates Z_∞ and $\beta \epsilon Z_2$. Therefore $h_{k+1}(k+1^{k+1})$ is the sub-group of Z_2+Z_∞ which is generated by $2\alpha+\beta$, and it follows that

$$h^{-1}_{k+1}(0) = 0.$$

Therefore $$\pi^2_{k3} = \phi_{k+2}(\pi^2_{k2}) = \phi_{k+2}(Z_2).$$

It follows from Theorem 9 that

$$\begin{aligned} h_{k+2}(\gamma) &= 2\alpha(\gamma)+\beta(\gamma) \\ &= \beta(\gamma), \end{aligned}$$

where γ is the non-zero element in $(k+1^{k+2})$ and α, β mean the same as before. Since $\beta \epsilon Z_2 = \phi_{k+1}(\pi^1_{k1})$ it is represented by a map of the form $g(S^{k+1}) \subset P^k_{k-1} = S^k$. Therefore $\beta(\gamma)$ is represented by the map

$$gg'(S^{k+2}) \subset S^k,$$

where $g'(S^{k+2}) \subset S^{k+1}$ is a map which represents γ. Therefore

$$\beta(\gamma) \epsilon (k^{k+2}) = 0,$$

whence $h_{k+2}(k+1^{k+2}) = 0$, and, by Theorem 8 (c), $\phi^{-1}_{k+2}(0) = 0$. Therefore

$$\pi^2_{k3} = Z_2 \quad (k = 2l+2).$$

Since the map $\phi u(S^{k+2}) = P^{k+2}_{k-1}$ is essential, according to Theorem 5, it represents the non-zero element in π^2_{k3}, whence

$$\pi^2_{km} = \pi^2_{k4} = 0 \quad (k = 2l+2;\ m \geqslant 4).$$

Now let $k = 3$ and let $w(V_{5,2}) = V_{5,1} = S^4$ be the map in which $w(v_1, v_2) = v_2$. Then $w^{-1}(a_2) = V_{4,1} = S^3$, and $(3^5) = 0$. Since $4 < 2k$ when $k = 3$ it follows, as in Theorem 3, that a given map

$$f_1(S^4) \subset S^3 \subset P_2{}^4$$

is essential in $V_{5,2}$ if it is essential in $P_2{}^4$. But we have proved† that

† Just after (6 . 2), in Theorem 10.

$\phi_4^{-1}(0) = 0$, where $\phi_4\{\pi_4(S^3)\} \subset \pi_4(P_2^4)$ is the homomorphism in which corresponding elements are represented by the same map. Therefore $f_1(S^4) \subset S^3$ is essential in P_2^4, and hence in $V_{5,2}$, if it is essential in S^3. It follows from the second corollary to Theorem 14, in § 7 below, that

$$\pi_5(V_{5,2}) \simeq \pi_5(V_{5,1}) = (4^5),$$

whence $$\pi_{32}^2 = Z_2.$$

Now let $k = 2l+1$ $(l \geqslant 1)$. Since $k \geqslant 3$ we have $\pi_s(P_{k-1}^{k+1}) = 0$ for $s = 1, 2$, and $k+2 < 2(k+2)-2$. Therefore if $r = t = k+2$, $X = P_{k-1}^{k+1}$, the conditions of Theorem 8 are satisfied. By Theorem 10,

$$\pi_{r-1}(X) = \pi_{k+1}(P_{k-1}^{k+1}) = \pi_{k2}^1 = Z_2.$$

Since h_{k+1} is a homomorphism of $(k+1^{k+1})$ in Z_2, we have

$$h_{k+1}^{-1}(0) = Z_\infty. \tag{6.3}$$

Since $k+1$ is even, $\phi u(S^{k+1}) \subset P_{k-1}^{k+1}$ is homotopic to a map $g(S^{k+1}) \subset P_{k-1}^k$. For $h_k^{-1}(0) = 0$, according to (6.1), whence it follows from Theorem 8 (a) that $\pi_{k2}^2 = \phi_{k+1}(\pi_{k1}^2)$. Therefore

$$h_{k+2}(k+1^{k+2}) = 0,$$

as when $k = 2l+2$. Therefore $\phi_{k+2}^{-1}(0) = 0$, by Theorem 8 (c), and it follows from Theorem 8 (a) that

$$\begin{aligned}\psi_{k+2}\{h_{k+1}^{-1}(0)\} &= \pi_{k3}^2 - \phi_{k+2}(\pi_{k2}^2) \\ &= \pi_{k3}^2 - Z_2.\end{aligned}$$

If $\alpha \in \pi_{k3}^2$ is either of the two elements in a residue class (mod Z_2), which generates $\psi_{k+2}\{h_{k+1}^{-1}(0)\}$, then any element in π_{k3}^2 can be expressed uniquely in the form $p\alpha + \beta$, where $\beta \in Z_2$. Therefore $\psi_{k+2}\{h_{k+1}^{-1}(0)\}$ may be identified with a sub-group of π_{k3}^2, and we have

$$\pi_{k3}^2 = Z_2 + Z_\infty \quad (k = 2l+1).$$

Now let $k = 2l+3$ $(l \geqslant 1)$. Since $\phi u(S^{k+1}) \subset P_{k-1}^{k+1}$ is essential we have $h_{k+1}(k+1^{k+1}) \neq 0$ $[h_{k+1}(k+1^{k+1}) \subset \pi_{k2}^1]$. Since π_{k2}^1 is of order two it follows that $h_{k+1}^{-1}(0) = 2(k+1^{k+1})$, and a generating element of $h_{k+1}^{-1}(0)$ is represented by a map $g(S^{k+1}) \subset S^{k+1}$, of degree 2. It follows from the definition of the homomorphism

$$\psi_{k+2}\{h_{k+1}^{-1}(0)\} \subset \pi_{k3}^2 - \phi_{k+2}(\pi_{k2}^2),$$

in Theorem 8, that the generator α of the summand Z_∞ in π_{k3}^2 is represented by a map $g(S^{k+2}) \subset P^{k+2}$, which covers P_{k-1}^{k+2} with degree ± 2. Also

an element $\beta \epsilon Z_2 = \phi_{k+2}(\pi^2_{k2})$ is represented by a map $g_1(S^{k+2}) \subset P^{k+2}_{k-1}$. Therefore a map which represents the element $p\alpha+\beta \epsilon \pi^2_{k3}$ has degree $\pm 2p$, and if a given map $g'(S^{k+2}) \subset P^{k+2}_{k-1}$ has degree ± 2, then it represents an element of the form $\pm\alpha+\beta$. Therefore the map $\phi u(S^{k+2}) \subset P^{k+2}_{k-1}$ represents an element of the form $\pm\alpha+\beta$, whence π^2_{k4} is given by generators α, β_1 subject to relations of the form

$$2\beta_1 = 0, \quad \epsilon\alpha+b\beta_1 = 0,$$

where $\epsilon = \pm 1$, $b = 0$ or 1. Whatever the values of ϵ and b, we have

$$\pi^2_{km} = \pi^2_{k4} = Z_2 \quad (k = 2l+3;\ m \geqslant 4).$$

Now let $k = 1$. Let $S^3 \subset R^4$ be given by

$$w\bar{w}+z\bar{z} = 1,$$

where $w = u+iv$, $z = x+iy$ are complex coordinates for R^4. The condition for two vectors whose complex coordinates are w, z and $w' = u'+iv'$, $z' = x'+iy'$ to be orthogonal is

$$\begin{aligned} 2(uu'+vv'+xx'+yy') &= w\bar{w}'+\bar{w}w'+z\bar{z}'+\bar{z}z' \\ &= 2Re(w\bar{w}'+z\bar{z}') = 0. \end{aligned}$$

We define a parallelism in S^3 by associating the vectors

$$X = (iw, iz), \quad Y = (-\bar{z}, \bar{w}), \quad Z = (-i\bar{z}, i\bar{w})$$

with a variable point $(w, z) \epsilon S^3$. It may be verified that X, Y, Z are mutually orthogonal, unit vectors, each of which is orthogonal to (w, z). Any vector which is orthogonal to (w, z) may be expressed in the form $sX+s'Y+s''Z$, where s, s', s'' are real parameters. Thus its complex coordinates (w', z') are given by

$$(6.4) \qquad \begin{cases} w' = isw-(s'+is'')\bar{z} = isw-\sigma\bar{z}, \\ z' = isz+(s'+is'')\bar{w} = isz+\sigma\bar{w}, \end{cases}$$

where $\sigma = s'+is''$. These equations can be solved to give

$$(6.5) \qquad \begin{cases} is = \bar{w}w'+\bar{z}z', \\ \sigma = -zw'+wz'. \end{cases}$$

If $sX+s'Y+s''Z$ is a unit vector we have

$$(6.6) \qquad s^2+\sigma\bar{\sigma} = 1,$$

and conversely. Thus we have an expression for $V_{4,m}$ ($m = 2$ or 3) as the topological product

$$V_{4,2} = S^3 \times V_{3,m-1},$$

where $V_{3,1}$ is given by (6.6) and $V_{3,2}$ consists of pairs of mutually orthogonal unit vectors (s_1, σ_1), (s_2, σ_2). Since $V_{3,2}$ is homeomorphic† to P^3, we have

$$\pi_{12}^2 = \pi_3(P^3) = Z_\infty,$$

$$\pi_{13}^2 = \pi_3(S^3 \times P^3) = Z_\infty + Z_\infty.$$

Let ξ_1, ξ_2, η_1, η_2 be real coordinates for P^3, such that

$$\xi_1^2 + \xi_2^2 + \eta_1^2 + \eta_2^2 = \xi\bar{\xi} + \eta\bar{\eta} = 1,$$

where $\xi = \xi_1 + i\xi_2$, $\eta = \eta_1 + i\eta_2$. As in (2.4), with slight modifications, the map $\phi(P^3) \subset V_{4,3}$ is given by

$$(u, v, x, y) = (-2\eta_2\xi_1, -2\eta_2\xi_2, -2\eta_2\eta_1, 1-2\eta_2^2),$$

$$(u_1, v_1, x_1, y_1) = (-2\eta_1\xi_1, -2\eta_1\xi_2, 1-2\eta_1^2, -2\eta_1\eta_2),$$

$$(u_2, v_2, x_2, y_2) = (1-2\xi_1^2, -2\xi_1\xi_2, -2\xi_1\eta_1, -2\xi_1\eta_2),$$

i.e. by

$$(6.7) \qquad \begin{cases} (a) & w = i(\eta - \bar{\eta})\,\xi, & z = i(\xi\bar{\xi} + \eta^2), \\ (b) & w_1 = -(\eta + \bar{\eta})\,\xi, & z_1 = \xi\bar{\xi} - \eta^2, \\ (c) & w_2 = \eta\bar{\eta} - \xi^2, & z_2 = -(\xi + \bar{\xi})\,\eta, \end{cases}$$

in which we have used the relation $\xi\bar{\xi} + \eta\bar{\eta} = 1$ to express z, z_1 and w_2. On substituting from (6.7) in (6.5), with $(w', z') = (w_j, z_j)$ $(j = 1, 2)$, we have

$$(6.8) \qquad \begin{cases} (a) & s_1 = \eta\bar{\eta} - \xi\bar{\xi}, & \sigma_1 = 2i\xi\eta, \\ (b) & s_2 = \xi\bar{\eta} + \bar{\xi}\eta, & \sigma_2 = i(\xi^2 - \eta^2). \end{cases}$$

From these equations we have

$$(6.9) \qquad \begin{cases} (a) & -s_1 + is_2 = (\xi + i\eta)(\bar{\xi} + i\bar{\eta}), \\ (b) & \sigma_1 - i\sigma_2 = (\xi + i\eta)^2, \quad \bar{\sigma}_1 - i\bar{\sigma}_2 = -(\bar{\xi} + i\bar{\eta})^2. \end{cases}$$

† This (well-known) fact follows from arguments given below.

Let s_j, σ_j $(j = 1, 2)$ be given (unit, orthogonal) vectors. Then

$$s_j{}^2+\sigma_j\bar{\sigma}_j = 1, \quad 2s_1s_2+\sigma_1\bar{\sigma}_2+\bar{\sigma}_1\sigma_2 = 0,$$

whence

$$(s_1-is_2)^2+(\sigma_1-i\sigma_2)(\bar{\sigma}_1-i\bar{\sigma}_2) = 0. \tag{6.10}$$

Let ξ_0, η_0 be solutions of the equations (6.9b), where the square roots are chosen so that, in consequence of (6.10), (6.9a) are also satisfied. It follows from (6.9b) that any other solutions of the combined equations (6.9) are of the form

$$\xi+i\eta = \epsilon_1(\xi_0+i\eta_0), \quad \xi-i\eta = \epsilon_2(\xi_0-i\eta_0),$$

where $\epsilon_j = \pm 1$. If either $\xi_0 \pm i\eta_0 = 0$, say $\xi_0-i\eta_0 = 0$, then we may take $\epsilon_1 = \pm 1$. Otherwise we have $\epsilon_1\epsilon_2 = 1$ from (6.9a). In either case the only solutions are (ξ_0, η_0) and $(-\xi_0, -\eta_0)$. Therefore it follows, *ab initio*, that $V_{3,2}$ is homeomorphic to P^3, and that the equations (6.8) represent a locally (1.1) map $\theta(S^3) \subset V_{3,2}$ if ξ, η are interpreted as coordinates for S^3. Therefore the map θ represents an element α which generates the (infinite, cyclic) group $\pi_3(V_{3,2})$, and $\pi_3(S^3 \times V_{3,2})$ $(= \pi_{13}^2)$ is freely† generated by α and a generating element $\beta \in \pi_3(S^3)$. The map

$$\phi u(S^3) \subset V_{4,3}$$

is given by (6.7), if ξ, η are interpreted as coordinates for S^3 instead of P^3. Therefore it represents the element $\alpha-2\beta \in \pi_{13}^2$. Since $V_{4,3}$ is 3-simple it follows from a theorem which I have proved elsewhere‡ that

$$\pi_{14}^2 = \pi_3(V_{4,3}+\mathcal{E}^4)$$

is generated by α, β subject to the relation $\alpha-2\beta$. Therefore

$$\pi_{1m}^2 = \pi_{14}^2 = Z_\infty \quad (m \geqslant 4).$$

Let $k = 2$, and let us represent $V_{4,2}$ as the product

$$V_{4,2} = S^3 \times S^2,$$

by means of the equations (6.4). Then we see that

$$\pi_{22}^2 = Z_2+Z_2.$$

Since $k = 2 > 1$ it follows from Theorem 2 that

$$\pi_{23}^2 = \pi_4(V_{4,2}+\mathcal{E}^4),$$

---

† Cf. Note 2 in § 8 below.

‡ *Proc. London Math. Soc.* (*loc. cit.*), Theorem 18. We cannot apply Theorem 8, since $\pi_1(V_{4,3}) \neq 0$.

and we can apply Theorem 8 (*a*), with $r = t = 4$, $X = V_{4,2}$, since $\pi_1(V_{4,2}) = 0$ and $r = 4 < 2t-2 = 6$. The map $\theta(S^3) \subset S^3$, which is given by (6.7*a*), with $(\xi, \eta) \in S^3$, has degree -2, and is therefore essential. Therefore the map $\phi u(S^3) \subset V_{4,2}$, which is given by (6.7*a*) and (6.7*b*), is essential, whence $h_3(3^3) \neq 0$. Since $\pi_3(V_{4,2})$ is a free Abelian group, namely $Z_\infty + Z_\infty$, and since (3^3) is a free cyclic group, it follows that $h_3^{-1}(0) = 0$, whence

$$\pi^2_{23} = \phi_4(\pi^2_{22}),$$

by reason of Theorem 8 (*a*). In calculating $\phi_4^{-1}(0)$ we use, not Theorem 8, since $r-t+2 = 2$ and $\pi_2(V_{4,2}) \neq 0$, but a lemma which is proved elsewhere†. According to the latter, $\phi_4^{-1}(0)$ is the group which is generated by the elements in $h_4(3^4)$, together with all elements of the form‡ $\alpha . \beta$, where $\alpha \in \pi_2(V_{4,2})$, $\beta \in h_3(3^3)$. Let us distinguish between the summands in π^2_{22} by writing $\pi^2_{22} = Z_2 + Z_2'$, where

$$Z_2 = (3^4), \quad Z_2' = (2^4).$$

Then I say that

$$\phi_4^{-1}(0) = Z_2'. \tag{6.11}$$

For let $\alpha - \beta$ [$\alpha \in (2^3)$, $\beta \in (3^3)$] be the element corresponding to the map $\phi u(S^3) \subset V_{4,2}$, which is given by (6.7*a*) and (6.7*b*), or by (6.7*a*) and (6.8*a*). On comparing (6.8*a*) with (4.1) it is clear that α generates (2^3). Also $\beta = 2\beta_1$, where β_1 generates (3^3), since the map given by (6.7*a*) is of degree -2. It follows from Theorem 9, with $r = 4$, $t = 3$, that $h_4(3^4)$ contains the element

$$\begin{aligned} \alpha(\gamma) + \beta(\gamma) &= \alpha(\gamma) - 2\beta_1(\gamma) \\ &= \alpha(\gamma) \in Z_2', \end{aligned}$$

where γ is the non-zero element in (3^4). Any map $\theta(S^4) \subset S^2$ is homotopic to a map of the form $\theta_2\theta_1$, where θ_1 and θ_2 map S^4 in S^3 and S^3 in S^2, respectively. Therefore any element $\alpha' \in (2^4)$ is of the form

$$\alpha' = p\alpha(q\gamma) = pq\alpha(\gamma) \quad (p = 0, \pm 1, \ldots;\ q = 0 \text{ or } 1),$$

since α, γ generate (2^3), (3^4). Since (2^4) is of order 2, it follows that $\alpha' = 0$ unless $p \equiv 1 \pmod 2$, $q = 1$, and, since $(2^4) \neq 0$, that $\alpha' \neq 0$ if $p = q = 1$.

† J. H. C. Whitehead, "On adding relations to homotopy groups", *Annals of Math.* 42 (1941), Lemma 4.

‡ *Ibid.* See also Note 2 in § 8 below.

Therefore $\alpha(\gamma) \neq 0$, whence

$$Z_2' = h_4(3^4) \subset \phi_4^{-1}(0).$$

Conversely, let

$$\alpha \in \pi_2(V_{4,2}) = \pi_2(S^3 \times S^2) = (2^2),$$

and let $\beta \in \pi_3(V_{4,2})$ be given by $\beta = \beta' + \beta''$, where $\beta' \in (2^3)$, $\beta'' \in (3^3)$. Since the transformation $\beta \to \alpha . \beta$ is a homomorphism, and since† $\alpha . \beta'' = 0$, we have

$$\alpha . (\beta' + \beta'') = \alpha . \beta' \in (2^4) = Z_2'.$$

Since $h_4(3^4) = Z_2'$, it follows that

$$\phi_4^{-1}(0) \subset Z_2',$$

whence $\phi_4^{-1}(0) = Z_2'$. Therefore π_{23}^2 is the residue group

$$\pi_{22}^2 - Z_2' = (Z_2 + Z_2') - Z_2',$$

whence

$$\pi_{23}^2 = Z_2.$$

Since‡ $\pi_4(V_{n,n-1}) = 0$ if $n \geqslant 6$ it follows from Theorem 12 below that

$$\pi_{2m}^2 = \pi_{24}^2 = 0 \quad (m \geqslant 4).$$

Let $r = t = r$, $X = S^3$, $X^* = P_2^4 = S^3 + \mathfrak{E}^4$ in Theorem 8. Then the conditions of Theorem 8 (a), namely $\pi_s(X) = 0$ $(1 \leqslant s \leqslant r - t + 1)$ and $r < 2t - 1$, are satisfied. Since $\phi u(S^3) \subset S^3$ is of degree -2, we have $h_3(3^3) = 2(3^3)$, whence $h_3^{-1}(0) = 0$. Therefore it follows from Theorem 8 (a) that any map $f_1^4(S^4) \subset P_2^4$ is homotopic to a map $g(S^4) \subset S^3$. Since $\phi u(S^4)$ is essential in P_2^4, a map in S^3 to which it is homotopic is essential in S^3. Since (3^4) is of order 2, it follows that any map $g(S^4) \subset S^3$, which is essential in S^3, is homotopic in P_2^4 to $\phi u(S^4)$. Therefore any map $g(S^4) \subset S^3$ is inessential in $P_2^5 = P_2^4 + \mathfrak{E}^5$, and hence in $V_{7,4}$. That is to say $\phi_4(3^4) = 0$, $\phi_4^{-1}(0) = (3^4)$, where $\phi_4(3^4) \subset \pi_{34}^1$ means the same as in (7.7). Hence, and from (7.7a), we have $\bar{\chi}_5(\pi_{43}^1) = (3^4)$, where $\bar{\chi}_5(\pi_{43}^1)$ means the same as in (7.7). But $\pi_{43}^1 = Z_2 + Z_2$, by Theorem 10, and is therefore of order 4, and since (3^4) is of order 2 it follows that $\bar{\chi}_5^{-1}(0)$ is of order $4/2 = 2$. Since $\pi_r(X_0) = 0$ when $r = 5$ and $X_0 = S^3$, it follows from (7.7b) and (7.7c) that

$$\pi_{3m}^2 = \pi_{34}^2 = Z_2 \quad (m \geqslant 4),$$

since $\phi u(S^4) \subset P_2^4$ is essential; and the proof is complete.

† See Note 2 in § 8 below.

‡ B. Eckmann, *loc. cit.*, Theorem VII, 2.

THEOREM 12. *If $r > 2$, $n > 2$, then*

$$\pi_r(G_n) = \pi_r(V_{n,n-2}),$$

where G_n is the group of rotations in R^n.

Let us identify G_n with $V_{n,n-1}$ and let $w(V_{n,n-1}) = V_{n,n-2}$ be the map in which $w(v_1, \ldots, v_{n-1}) = (v_2, \ldots, v_{n-1})$. Then

$$w^{-1}(a_2, \ldots, a_{n-1}) = V_{2,1} = S^1,$$

and since $\pi_{r-1}(S^1) = \pi_r(S^1) = 0 \quad (r > 2)$, Theorem 12 follows from Theorem 14, below, Corollary 2.

We now show that $\phi u(S^4) \subset P_2^4$ is essential, thus completing the proof of Theorem 5. By Theorem 12 and Eckmann's theorem,

$$\pi_4(V_{6,4}) = \pi_4(G_6) = 0,$$

and, by Theorem 2,

$$\pi_4(V_{6,4}) = \pi_4(V_{5,3} + \mathfrak{E}^5).$$

It follows from Theorem 8, with $r = 4$, $t = 5$, $X = V_{5,3}$ that if $\phi u(S^4) \subset V_{5,3}$ were inessential in $V_{5,3}$, then

$$\pi_4(V_{5,3} + \mathfrak{E}^5) = \pi_4(V_{5,3}) = \pi_{23}^2.$$

But in Theorem 11 it was proved that $\pi_{23}^2 = Z_2$, without assuming that $\phi u(S^4) \subset P_2^4$ is essential. Therefore $\phi u(S^4) \subset V_{5,3}$ is essential in $V_{5,3}$.

Now let $w(V_{5,3}) = V_{5,2}$ be the map which is given by

$$w(v_1, v_2, v_3) = (v_2, v_3),$$

and let us denote the map $\phi u(S^4) \subset P_1^4 \subset V_{5,3}$ by $\phi' u$, to distinguish it from $\phi u(S^4) \subset P_2^4 \subset V_{5,2}$. Assume that $\phi u(S^4) \subset P_2^4$ is inessential in P_2^4, and hence in $V_{5,2}$. Then it follows from Lemma 4, below, that $\phi' u(S^4) \subset V_{5,3}$ is homotopic to a map $g(S^4) \subset w^{-1}(a_2, a_3) = V_{3,1} = S^2$. When $V_{4,2}$ is represented as the product $S^3 \times S^2$, by means of the equations (6.5), we can obviously take $V_{3,1} = a_2 \times S^2$. Hence it follows from (6.11) that any map $g(S^4) \subset V_{3,1}$ is inessential in $V_{5,3}$. Therefore, if $\phi u(S^4)$ is inessential in P_2^4, so is $\phi' u(S^4)$ in $V_{5,3}$. But $\phi' u(S^4)$ is essential in $V_{5,3}$ and so, therefore, is $\phi u(S^4) \subset P_2^4$.

7. In this section we prove a theorem to which we have already referred more than once. The theorem can best be stated in terms of

relative homotopy groups†, concerning which we need certain elementary theorems. First of all we prove a lemma. Let E^r be an r-element, which is a sub-complex of a simplicial complex K. Let X be an arbitrary topological space and let $f_i(K) \subset X$ $(i = 0, 1)$ be two maps such that $f_i(E^r) = x_0$, where x_0 is a point in X. Then the lemma is:

LEMMA 3. *If there is a deformation $f_t(K) \subset X$ $(0 \leqslant t \leqslant 1)$, such that $f_t(p_0) = x_0$, where p_0 is an arbitrary point in E^r, then there is a deformation $f_t'(K) \subset X$ $(f_0' = f_0, f_1' = f_1)$, such that $f_t'(E^r) = x_0$.*

Let $g(K \times \langle 0, 1\rangle) \subset X$ be the map which is defined by $g(p \times t) = f_t(p)$ $(0 \leqslant t \leqslant 1)$. Then $g(F) = x_0$, where

$$F = (E^r \times 0) + (p_0 \times \langle 0, 1\rangle) + (E^r \times 1).$$

Clearly F is a retract by deformation of $E^r \times \langle 0, 1\rangle$, that is to say, there is a deformation $\delta_\lambda(E^r \times \langle 0, 1\rangle) \subset E^r \times \langle 0, 1\rangle$ $(0 \leqslant \lambda \leqslant 1)$, such that $\delta_0(p) = p$, $\delta_1(p) \in F$, $\delta_1(q) = q$, for any points $p \in E^r \times \langle 0, 1\rangle$, $q \in F$. Let us extend δ_λ throughout

$$(K \times 0) + (E^r \times \langle 0, 1\rangle) + (K \times 1)$$

by defining $\delta_\lambda(p) = p$ for every point $p \in (K \times 0) + (K \times 1)$. Let the deformation

$$g_\lambda\{(K \times 0) + (E^r \times \langle 0, 1\rangle) + (K \times 1)\} \subset X,$$

where $g_\lambda = g\delta_\lambda$, be extended throughout‡ $K \times \langle 0, 1\rangle$. Then $g_1 = g$ in $(K \times 0) + (K \times 1)$ and

$$\begin{aligned} g_1(E^r \times \langle 0, 1\rangle) &= g\delta_1(E^r \times \langle 0, 1\rangle) \\ &= g(F) \\ &= x_0. \end{aligned}$$

Therefore $f_t'(K)$, given by $f_t'(p) = g_1(p \times t)$, is a deformation of f_0 into f_1, such that $f_t'(E^r) = x_0$, and the lemma is established.

Now let X and $X_0 \subset X$ be arcwise connected topological spaces, and let $E^r \subset R^r$ $(r \geqslant 2)$ be the r-element which is given by $y_1^2 + \ldots + y_r^2 \leqslant 1$, in rectangular Cartesian coordinates $y_1, \ldots, y_r$ for R^r. An element in the

† Cf. a paper by W. Hurewicz and N. E. Steenrod, *Proc. Nat. Academy of Sciences*, 27 (1941), 60–64, which is closely related to this section.

‡ Alexandroff and Hopf, *loc. cit.*, p. 501, Lemma 1a.

relative homotopy group $\rho_r(X\,;\,X_0)$ is represented by a map $f(E^r)\subset X$, such that

$$(7.1)\qquad \begin{cases} (1)\quad f(\dot{E}^r)\subset X_0, \\ (2)\quad f(p_0)=x_0, \end{cases}$$

where $p_0=(1,\,0,\,\ldots,\,0)$, and x_0 is a point in X_0, which is chosen as the base point for $\rho_r(X\,;\,X_0)$. Two such maps, f_0 and f_1, represent the same element in $\rho_r(X\,;\,X_0)$ if, and only if, there exists a deformation f_t $(0\leqslant t\leqslant 1)$, such that

$$(1)\quad f_t(\dot{E}^r)\subset X_0,$$

$$(2)\quad f_t(p_0)=x_0.$$

Let E_1^r be the part of E^r for which $y_r\geqslant 0$ and E_2^r the part for which $y_r\leqslant 0$. A given element in $\rho_r(X\,;\,X_0)$ can obviously be represented by a map $f(E^r)\subset X$, such that $f(E_i^r)=x_0$ $(i=1$ or $2)$. Let given elements $\alpha_1,\,\alpha_2\subset\rho_r(X\,;\,X_0)$ be represented by maps f_1, f_2, such that $f_i(E_j^r)=x_0$ $(i,j=1,\,2\,;\,j\neq i)$. Then it follows from Lemma 3 that the element α, which is represented by the map $f(E^r)$ such that $f=f_i$ in E_i^r, depends only on the elements $\alpha_1,\,\alpha_2$. We write $\alpha=\alpha_1+\alpha_2$. The zero element 0 is represented by the map† $f(E^r)=x_0$, and $-\alpha$ is represented by $ft(E^r)$, where $f(E^r)$ represents α, and $t(p)$ is the mirror image of p in the hyperplane $y_r=0$. Thus $-(-\alpha)=\alpha$, and it is easily verified that

$$\alpha+0=\alpha,\quad \alpha+(-\alpha)=0,\quad \alpha_1+(\alpha_2+\alpha_3)=(\alpha_1+\alpha_2)+\alpha_3,$$

whence $\rho_r(X\,;\,X_0)$ is indeed a group. If $r>2$, then E_1^r and E_2^r can be interchanged by a rotation about the flat $(r-2)$-space which is given by $y_{r-1}=y_r=0$, and it follows that $\rho_r(X\,;\,X_0)$ is commutative‡. Notice that, if X_0 is a single point, then $\rho_r(X\,;\,X_0)$ is the homotopy group $\pi_r(X)$, with X_0 as base point; also that, if Y is any arcwise connected topological space, then a map $f(X)\subset Y$ induces a homomorphism of $\rho_r(X\,;\,X_0)$ in $\rho_r(Y\,;\,Y_0)$, where $Y_0=f(X_0)$.

† It follows that $0\in\rho_r(X\,;\,X_0)$ is represented by any map $f(E)\subset X_0$, such that $f(p_0)=x_0$.

‡ If X is a simplicial complex and if X_0 is a sub-complex of X, then $\rho_2(X\,;\,X_0)$ may be expressed in terms of the group h_Γ, which is defined in § 6 of the paper "On adding relations to homotopy groups" (*Annals of Math., loc. cit.*). In general it is non-commutative. For example, if X consists of two 2-elements, with a single point in common, and if X_0 consists of their boundaries, then $\rho_2(X\,;\,X_0)$ is a free group with two free generators. However, we shall write $\rho_2(X\,;\,X_0)$ as an additive group.

Let E_0^r $(r \geqslant 2)$ be the part of E^r in which $y_1 \geqslant 0$, and let a given element $\alpha \in \rho_r(X;\ X_0)$ be represented by a map $f(E^r)$ such that $f(E_0^r) = x_0$. Let $x = x(t)$ $[0 \leqslant t \leqslant 1,\ x(0) = x(1) = x_0,\ x(t) \in X_0]$ be an oriented circuit which represents a given element $\xi \in \pi_1(X_0)$, the positive sense being the one in which t increases. Let $\alpha_1 \in \rho_r(X;\ X_0)$ be the element which is represented by the map $f'(E^r)$, such that $f' = f$ in $E^r - E_0^r$ and

$$f'(E_t^{r-1}) = x(1-t) \quad (0 \leqslant t \leqslant 1),$$

where E_t^{r-1} is the intersection of E^r with the hyperplane $y_1 = t$. It is obvious that α_1 does not depend on the particular circuit $x = x(t)$, by which ξ is represented, and it follows from Lemma 3 that it does not depend on the particular map f which represents α. Therefore it depends only on ξ and α, and we write

$$\alpha_1 = \xi\alpha.$$

Given $\xi,\ \eta \subset \pi_1(X_0)$ and $\alpha \in \rho_r(X;\ X_0)$, it is obvious that

$$\xi(\eta\alpha) = (\xi\eta)\,\alpha.$$

Let given elements $\alpha_1,\ \alpha_2 \subset \rho_r(X;\ X_0)$ be represented by maps f_1, f_2 such that $f_j(y_1, \ldots, y_r) = x_0$ if $y_1 \geqslant 0$, and also

$$f_1(y_1, \ldots, y_r) = x_0 \quad \text{if} \quad y_r \leqslant 0,$$

$$f_2(y_1, \ldots, y_r) = x_0 \quad \text{if} \quad y_r \geqslant 0,$$

and let $\alpha_1 + \alpha_2$ be given by the map f such that

$$f(y_1, \ldots, y_r) = f_1(y_1, \ldots, y_r) \quad \text{if} \quad y_r \geqslant 0$$

$$= f_2(y_1, \ldots, y_r) \quad \text{if} \quad y_r \leqslant 0.$$

Then, given $\xi \in \pi_1(X_0)$, it is obvious from the above definition of $\xi\alpha$ that

$$\xi(\alpha_1 + \alpha_2) = \xi\alpha_1 + \xi\alpha_2.$$

Clearly $\xi\alpha = \alpha$ if $\xi = 1$, and $\xi\alpha = 0$ if $\alpha = 0$. Therefore $\alpha \to \xi\alpha$ is an automorphism ψ_ξ of $\rho_r(X;\ X_0)$, and $\xi \to \psi_\xi$ is a homomorphism of $\pi_1(X_0)$ in the group of automorphisms of $\rho_r(X;\ X_0)$. If $r > 2$, then we define

$$(\xi - \eta)\alpha = \xi\alpha - \eta\alpha,$$

and so exhibit $\rho_r(X;\ X_0)$ as a group with operators in the group ring of $\pi_1(X_0)$.

Let x_0 be chosen as the base point for $\pi_{r-1}(X_0)$, let $f(E^r)$ be a map which represents a given element $\alpha \in \rho_r(X;\ X_0)$, and let $\beta \in \pi_{r-1}(X_0)$ be

the element which is represented by the map $f(\dot{E}^r) \subset X_0$. It is obvious that β depends only on the element α, and we therefore write

$$\beta = \chi_r(\alpha).$$

Given $\alpha_1, \alpha_2 \subset \rho_r(X;\ X_0)$, $\xi \in \pi_1(X_0)$, it is also obvious that

$$\chi_r(\alpha_1+\alpha_2) = \chi_r(\alpha_1)+\chi_r(\alpha_2),$$

and that

$$\chi_r(\xi\alpha) = \xi\chi_r(\alpha)\,\xi^{-1} \quad \text{if} \quad r=2$$
$$= \xi\chi_r(\alpha) \qquad \text{if} \quad r>2;$$

also that

$$\chi_r\{\rho_r(X;\ X_0)\} = \phi_{r-1}^{-1}(0), \tag{7.2}$$

where $\phi_s\{\pi_s(X_0)\} \subset \pi_s(X)$ $(s = 1, 2, \ldots)$ is the homomorphism in which corresponding elements are represented by the same map.

A given element $\gamma \in \pi_r(X)$, with x_0 as base point, is represented by a map $f(E^r) \subset X$, with $f(\dot{E}^r) = x_0$, and this map also represents an element $\alpha \in \rho_r(X;\ X_0)$. Clearly α depends only on γ, and we write

$$\alpha = \psi_r(\gamma).$$

It is also obvious that ψ_r is a homomorphism of $\pi_r(X)$ in $\rho_r(X;\ X_0)$, and that

$$\psi_r\{\pi_r(X)\} = \chi_r^{-1}(0). \tag{7.3}$$

If $\gamma \in \phi_r\{\pi_r(X_0)\}$, there is a deformation f_t $(0 \leqslant t \leqslant 1)$, of a map $f_0(E^r) \subset X$ $[f_0(\dot{E}^r) = x_0]$, which represents γ, such that $f_t(\dot{E}^r) = x_0$ and $f_1(E^r) \subset X_0$, whence $\psi_r(\gamma) = 0$. Conversely, let $f_0(E^r) \subset X$ $[f_0(\dot{E}^r) = x_0]$ represent a given element $\gamma \in \psi_r^{-1}(0)$. Since $\psi_r(\gamma) = 0$ there is a deformation f_t, which satisfies the conditions (7.1), such that $f_1(E^r) = x_0$. Let $S_\lambda^{r-1} \subset E^r$ $(0 \leqslant \lambda \leqslant 1)$ be the locus $y_1^2+\ldots+y_r^2 = \lambda$, and let f_t' $(0 \leqslant t \leqslant 1;\ f_0' = f_0)$ be the deformation of f_0, which is given by

$$f_t'(S_\lambda^{r-1}) = f_t\{S_{(1+t)\lambda}^{r-1}\} \qquad \text{if} \quad \lambda \leqslant \tfrac{1}{2}$$
$$= f_{2t(1-\lambda)}\{S_{\lambda+(1-\lambda)t}^{r-1}\} \quad \text{if} \quad \lambda \geqslant \tfrac{1}{2}.$$

Then

$$f_1'(S_\lambda^{r-1}) = f_1(S_{2\lambda}^{r-1}) = x_0 \qquad \text{if} \quad \lambda \leqslant \tfrac{1}{2}$$
$$= f_{2(1-\lambda)}(S_1^{r-1}) \subset X_0 \quad \text{if} \quad \lambda \geqslant \tfrac{1}{2},$$

and
$$f_t'(\dot{E}^r) = f_t'(S_1^{r-1})$$
$$= f_0(S_1^{r-1}) = x_0.$$

Therefore the map f_0 is deformable, with $f_0(\dot{E}^r) = x_0$ held fixed, into the map $f_1'(E^r) \subset X_0$. Therefore $\gamma \in \phi_r\{\pi_r(X_0)\}$, and we have

$$\phi_r\{\pi_r(X_0)\} = \psi_r^{-1}(0). \tag{7.4}$$

Collecting these results, we have the theorem:

THEOREM 13. *The groups* $\pi_s(X_0)$, $\pi_s(X)$ $(s = r-1,\ r;\ r \geqslant 2)$ *and* $\rho_r(X;\ X_0)$ *are related by homomorphisms*

$$\phi_s\{\pi_s(X_0)\} \subset \pi_s(X), \quad \chi_r\{\rho_r(X;\ X_0)\} \subset \pi_{r-1}(X_0) \quad \text{and} \quad \psi_r\{\pi_r(X)\} \subset \rho_r(X;\ X_0),$$

such that

$$\begin{cases} (a) & \chi_r\{\rho_r(X;\ X_0)\} = \phi_{r-1}^{-1}(0), \\ (b) & \psi_r\{\pi_r(X)\} = \chi_r^{-1}(0), \\ (c) & \phi_r\{\pi_r(X_0)\} = \psi_r^{-1}(0), \end{cases} \tag{7.5}$$

and $\chi_r(\xi\alpha) = \xi\chi_r(\alpha)\,\xi^{-1}$ *or* $\xi\alpha$, *according as* $r = 2$ *or* $r > 2$, *where* $\xi \in \pi_1(X_0)$, $\alpha \in \rho_r(X;\ X_0)$.

Let T be an arbitrary topological space, and let a given topological space X be a fibre space over a base space Y, the fibres being homeomorphic to T. That is to say, there is a map $w(X) \subset Y$, such that

(F) *there is a basis* $\{U\}$, *for the open sets in* Y, *to each member of which corresponds a homeomorphism*

$$\psi(U \times T) = w^{-1}(U) \subset X,$$

such that

$$\psi(y \times T) = w^{-1}(y)$$

for each $y \in U$.

Let K be a simplicial complex and $L \subset K$ a sub-complex which is a retract by deformation of K. Let $f_0(L) \subset X$ be a given map and let

$$g_0(L) = wf_0(L) \subset Y.$$

LEMMA 4. *Let* g_0 *be extended to a map* $g(K) \subset Y$. *Then* f_0 *can be extended to a map* $f(K) \subset X$, *such that* $g = wf$.

First let

$$K = A \times \langle 0, 1\rangle, \quad L = (A \times 0) + (\dot{A} \times \langle 0, 1\rangle),$$

where A is a (closed) simplex†, and let there be a subdivision of $\langle 0, 1\rangle$, whose vertices are $0 = s_0 < s_1 < \ldots < s_h = 1$, such that

$$g(A \times \langle s_i, s_{i+1}\rangle) \subset U_i \quad (i = 0, \ldots, h-1),$$

where $U_i \in \{U\}$. Assume that f_0 has been extended to a map

$$f_i(A \times \langle 0, s_i\rangle + \dot{A} \times \langle 0, 1\rangle) \subset X \quad (0 \leqslant i < h),$$

such that $g = wf_i$ in $A \times \langle 0, s_i\rangle + \dot{A} \times \langle 0, 1\rangle$. Then

$$f_i(A_i + C_i) \subset w^{-1}(U_i) = \psi_i(U_i \times T),$$

where $A_i = A \times s_i$, $C_i = \dot{A} \times \langle s_i, s_{i+1}\rangle$ and ψ_i is a homeomorphism which, with U_i, satisfies the condition (F) above. Let us write

$$\psi_i^{-1} f_i(A_i + C_i) = y(A_i + C_i) \times t(A_i + C_i),$$

where $y \in U_i$, $t \in T$. Since $w\psi_i(y \times T) = y$, it follows that, if $p \in A_i + C_i$, then

$$\begin{aligned} y(p) &= w\psi_i\{y(p) \times t(p)\} \\ &= wf_i(p) \\ &= g(p). \end{aligned}$$

Therefore the map $y(A_i + C_i)$ can be extended to the map

$$y(A \times \langle s_i, s_{i+1}\rangle) = g(A \times \langle s_i, s_{i+1}\rangle).$$

Clearly $A_i + C_i$ is a retract of $A \times \langle s_i, s_{i+1}\rangle$, say $\theta(A \times \langle s_i, s_{i+1}\rangle) = A_i + C_i$, where $\theta(p) = p$ if $p \in A_i + C_i$. Therefore $t(A_i + C_i)$ can be extended to the map $t\theta(A \times \langle s_i, s_{i+1}\rangle) \subset T$. Therefore $f_i(A \times \langle 0, s_i\rangle + \dot{A} \times \langle 0, 1\rangle)$ can be extended to the map $f_{i+1}(A \times \langle 0, s_{i+1}\rangle + \dot{A} \times \langle 0, 1\rangle) \subset X$, which is given by

$$\begin{aligned} f_{i+1}(p) &= f_i(p) && \text{if} \quad p \in A \times \langle 0, s_i\rangle + \dot{A} \times \langle 0, 1\rangle \\ &= \psi_i\{g(p) \times t\theta(p) && \text{if} \quad p \in A \times \langle s_i, s_{i+1}\rangle. \end{aligned}$$

Since $w\psi_i\{g(p) \times t\theta(p)\} = g(p)$, we have $wf_{i+1} = g$ in $A \times \langle 0, s_{i+1}\rangle + \dot{A} + \langle 0, 1\rangle$, and this special case of the lemma follows from induction on i.

† A may be a 0-simplex, in which case $\dot{A}$ is the empty set.

Now let K be any simplicial complex and let $L \subset K$ be a sub-complex which is a retract by deformation of K. That is to say, there is a map

$$\theta(K \times \langle 0, 1\rangle) \subset K,$$

such that $\theta(p \times 0) = p$, $\theta(K \times 1) = L$, $\theta(q \times 1) = q$, where $p \in K$, $q \in L$. Moreover we may assume that† $\theta(q \times t) = q$ $(0 \leqslant t \leqslant 1;\ q \in L)$. This being so, let

$$L' = (K \times 1) + (L \times \langle 0, 1\rangle),$$

and let $f_0'(L') \subset X$, $g_0'(L') \subset Y$ and $g'(K \times \langle 0, 1\rangle) \subset Y$ be given by $f_0' = f_0\theta$, $g_0' = g_0\theta$ and $g' = g\theta$. Then $g_0' = g_0\theta = wf_0\theta = wf'_0$. After a suitable subdivision of K we assume that there is a subdivision of $\langle 0, 1\rangle$, with end points $0 = s_0 < s_1 < \ldots < s_h = 1$, such that $g'(A \times \langle s_i, s_{i+1}\rangle)$ is in one of the sets $\{U\}$, for each simplex A in K, and each‡ $i = 0, \ldots, h-1$. Let $K_r = K^r + L$ $(r \geqslant -1,\ K_{-1} = L)$, let

$$L_r' = (K \times 1) + (K_r \times \langle 0, 1\rangle) \quad (L'_{-1} = L)$$

and assume that f_0' has been extended to a map $f'_{r+1}(L_r') \subset X$, such that $g' = wf'_{r+1}$ in L_r'. If $r < \dim(K)$, it follows from what we have already proved that f'_{r+1} can be extended throughout $A^{r+1} \times \langle 0, 1\rangle$, for each $(r+1)$-simplex whose interior is in $K - L$, to a map $f'_{r+2}(L'_{r+1}) \subset X$, such that $g' = wf'_{r+2}$ in L'_{r+1}. It follows from induction on r that f_0' can be extended to a map $f'(K \times \langle 0, 1\rangle) \subset X$, such that $g' = wf'$. Let $f(K) \subset X$ be given by

$$f(p) = f'(p \times 1).$$

Then $f(q) = f_0'(q \times 1) = f_0(q)$ $(q \in L)$,

$$g(p) = g\theta(p \times 1) = g'(p \times 1) = wf'(p \times 1) = wf(p),$$

and the lemma is established.

Now let X, and hence Y, and also let T be arcwise connected. Let $X_0 = w^{-1}(y_0)$, where y_0 is an arbitrary point in Y. Then w induces a homomorphism of $\rho_r(X;\ X_0)$ in $\rho_r(Y;\ y_0) = \pi_r(Y)$, which we denote by θ_r $(r \geqslant 2)$.

THEOREM 14. *The homomorphism θ_r is an isomorphism of $\rho_r(X;\ X_0)$ on $\pi_r(Y)$.*

† Whitehead, *Proc. London Math. Soc.* (*loc. cit.*), Theorem 15, p. 273.

‡ If K is finite this assumption is justified by the Heine-Borel theorem, and if K is infinite by Whitehead, *Proc. London Math. Soc.* (*loc. cit.*), Theorem 35, p. 317. We need the lemma only when K is finite.

Let y_0 be taken as the base point for $\pi_r(Y)$. Then an arbitrary element $\beta \in \pi_r(Y)$, is represented by a map $g(E^r) \subset Y$, such that $g(\dot{E}^r) = y_0$, where E^r means the same as before. Since p_0 is a retract by deformation of E^r, where $p_0 \in \dot{E}^r$ means the same as before, the map $f_0(p_0) = x_0 \in X_0$ can be extended to a map $f(E^r) \subset X$, such that $g = wf$, according to Lemma 4. Then $f(\dot{E}^r) \subset w^{-1}(y_0) = X_0$, whence $\beta = \theta_r(\alpha)$, where $\alpha \in \rho_r(X\,;\, X_0)$ is represented by $f(E^r)$. Therefore θ_r is a homomorphism of $\rho_r(X\,;\, X_0)$ on $\pi_r(Y)$. Let $f_0(E^r)$ be a map which represents a given element $\alpha \in \theta_r^{-1}(0)$, let $g_0 = wf_0$, and let us imbed E^r in $E^r \times \langle 0, 1\rangle$ by identifying $p \in E^r$ with $p \times 0$. Since $\theta_r(\alpha) = 0$, the map $g_0(E^r \times 0)$ can be extended to a map $g(E^r \times \langle 0, 1\rangle) \subset Y$, such that

$$g(\dot{E}^r \times \langle 0, 1\rangle) = y_0, \quad g(E^r \times 1) = y_0.$$

We extend f_0 throughout $(E^r \times 0) + (p_0 \times \langle 0, 1\rangle)$ by defining

$$f_0(p_0 \times \langle 0, 1\rangle) = x_0.$$

Clearly $(E^r \times 0) + (p_0 \times \langle 0, 1\rangle)$ is a retract by deformation of $E^r \times \langle 0, 1\rangle$. Therefore it follows from Lemma 4 that f_0 can be extended to a map $f(E^r \times \langle 0, 1\rangle) \subset X$, such that $g = wf$. Since $g = wf$, we have

$$f(\dot{E}^r \times \langle 0, 1\rangle) \subset X_0,$$

and $f(E^r \times 1) \subset X_0$. Therefore $f(E^r)$ is homotopic, subject to the conditions (7.1), to a map $f_1(E^r) \subset X_0$, and hence to the map $f_2(E^r) = x_0$. Therefore $\alpha = 0$, and the theorem is established.

If $h_r\{\pi_r(X)\} \subset \pi_r(Y)$ is the homomorphism which is induced by the map w, then $h_r = \theta_r \psi_r$, where ψ_r means the same as in Theorem 13. Therefore, writing $\bar{\chi}_r = \chi_r \theta_r^{-1}$, we have the corollary to Theorems 13 and 14:

COROLLARY 1†. *There is a homomorphism* $\bar{\chi}_r\{\pi_r(Y)\} \subset \pi_{r-1}(X_0)$, *which is in the following relationship to the homomorphisms* ϕ_s $(s = r-1, r)$ *and* h_r:

$$(7.7) \qquad \begin{cases} (a) & \bar{\chi}_r\{\pi_r(Y)\} = \phi_{r-1}^{-1}(0), \\ (b) & h_r\{\pi_r(X)\} = \bar{\chi}_r^{-1}(0), \\ (c) & \phi_r\{\pi_r(X_0)\} = h_r^{-1}(0). \end{cases}$$

If any map $f(S^{r-1}) \subset X_0$, which is essential in X_0, is essential‡ in X, then $\phi_{r-1}^{-1}(0) = 0$, and it follows from (7.7a) that $\bar{\chi}_r\{\pi_r(Y)\} = 0$, whence

† Cf. W. Hurewicz, *Proc. Akad. Amsterdam*, 38 (1935), 112–119, Theorem 12.

‡ *E.g.* if $\pi_{r-1}(X_0) = 0$ (or 1 if $r = 2$).

$\chi_r^{-1}(0) = \pi_r(Y)$. Therefore h_r determines an isomorphism of the residue group $\pi_r(X) - \phi_r\{\pi_r(X_0)\}$ on $\pi_r(Y)$, according to (7.7*b*) and (7.7*c*). Therefore we have

COROLLARY 2. *If any map $f(S^{r-1}) \subset X_0$, which is essential in X_0, is essential in X, then*

$$\pi_r(X) - \phi_r\{\pi_r(X_0)\} \simeq \pi_r(Y).$$

If $\alpha \in \rho_r(X; X_0)$, $\xi \in \pi_1(X_0)$, it is obvious from the definition of $\xi\alpha$ that $\theta_r(\xi\alpha) = \theta_r(\alpha)$, and it follows from Theorem 14 that $\xi\alpha = \alpha$. Therefore $\chi_r(\xi\alpha) = \chi_r(\alpha)$ and, with the help of Theorem 13, we have

COROLLARY 3. *The group $\phi_1^{-1}(1)$ is contained in the centre of $\pi_1(X_0)$. If $r > 2$, then each element in $\phi_{r-1}^{-1}(0)$ is invariant under the operators in the group ring of $\pi_1(X_0)$.*

8. *Note* 1. Let K and L be simplicial complexes, let K_0 be a subcomplex of K and let $f_0(K) \subset L$ be a map which is simplicial in K_0. Let $s_0 K$ be the complex which is derived from K by starring, in order of decreasing dimensionality, the simplexes in K, whose interiors are in $K - K_0$, and let $s_0^{p+1} K = s_0(s_0^p K)$ for $p = 1, 2, \ldots$.

THEOREM 15. *There is a deformation f_t $(0 \leqslant t \leqslant 1)$ of f_0 into a map f_1, which is simplicial with respect to $s_0^p K$ for a suitable value of p, such that $f_t(p) = p$ if $p \in K_0$.*

If f_0 is simplicial, not merely in K_0, but in the simplicial neighbourhood $N(K_0) \subset K$ of K_0, then the theorem follows from standard arguments, with trivial modifications. Otherwise, let $\theta_t(s_0 K)$ $(0 \leqslant t \leqslant \alpha < 1)$ be a canonical deformation of $s_0 K$ into K, such that $\theta_\alpha\{N'(K_0)\} \subset K_0$, where $N'(K_0)$ is the simplicial neighbourhood of K_0 in $s_0 K$. Then $f_t = f_0 \theta_t$ is a deformation of f_0 into a map $f_\alpha(s_0 K)$, which is simplicial in $N'(K_0)$, and the theorem follows.

Note 2. *Concerning* $\pi_r(X_1 \times X_2)$. The group $\pi_r(X_1 \times X_2)$ $(r \geqslant 1)$ is isomorphic to the direct sum $\pi_r(X_1) + \pi_r(X_2)$, or to their direct product if $r = 1$ and the groups are multiplicative, where X_1 and X_2 are arcwise connected topological spaces. For let $\alpha \in \pi_r(X_1 \times X_2)$ be the element which is represented by a given map $f(S^r) \subset X_1 \times X_2$, with base point $b_1 \times b_2$ $(b_i \in X_i)$. Let $f(p) = f_1(p) \times f_2(p)$ $(p \in S^r)$ and let $\beta_i \in \pi_r(X_i)$ $(i = 1, 2)$ be the element which is represented by the map $f_i(S^r) \subset X_i$, with b_i as

base point. Then it is easily verified that the transformation $\psi(\alpha) = \beta_1 + \beta_2$, or $\beta_1 \times \beta_2$ if $r = 1$, is an isomorphism of $\pi_r(X_1 \times X_2)$ on $\pi_r(X_1) + \pi_r(X_2)$, or on $\pi_r(X_1) \times \pi_r(X_2)$ if $r = 1$. The relation $\beta_1\beta_2 = \beta_2\beta_1$, when $r = 1$, has the following generalization. Let

$$\beta_1 \in \pi_r(X_1 \times X_2), \quad \beta_2 \in \pi_s(X_1 \times X_2)$$

be elements which are represented by

$$g_1(E_1^r) = f_1(E_1^r) \times b_2, \quad g_2(E_2^s) = b_1 \times f_2(E_2^s),$$

where $f_1(E_1^r) \subset X_1$, $f_2(E_2^s) \subset X_2$ are given maps such that $f_1(\dot{E}_1^r) = b_1$, $f_2(\dot{E}_2^s) = b_2$. Then $\beta_1 . \beta_2$ is the zero element in $\pi_{r+s-1}(X_1 \times X_2)$. For, neglecting orientation, $\beta_1 . \beta_2$ corresponds to the map

$$f(\dot{E}_1^r \times E_2^s + E_1^r \times \dot{E}_2^s) \subset X_1 \times X_2,$$

where

$$f(\dot{E}_1^r \times p_2) = b_1 \times f_2(p_2),$$

$$f(p_1 \times \dot{E}_2^s) = f_1(p_1) \times b_2 \quad (p_1 \in E_1^r, \ p_2 \in E_2^s),$$

and f can be extended to the map $f(E_1^r \times E_2^s) \subset X_1 \times X_2$, where

$$f(p_1 \times p_2) = f_1(p_1) \times f_2(p_2).$$

This is the reason, or at least a reason, why $\pi_t(X_1 \times X_2)$ is isomorphic to the direct sum, or product, of $\pi_t(X_1)$ and $\pi_t(X_2)$, while $\pi_t(X_1 + X_2)$ in general is not, where $X_1 + X_2$ consists of the spaces X_1 and X_2 with a single point in common.

Note 3. Our results have an entertaining application to the topology of a non-degenerate quadric Q, in projective space of $n-1$ complex dimensions ($n \geqslant 3$). For we have

THEOREM 16. *If $r > 2$, then $\pi_r(Q) = \pi_r(V_{n,2})$.*

A point in $V_{n,2}$ consists of points $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \subset R^n$, subject to the conditions

$$\Sigma x_j^2 = \Sigma y_j^2 = 1, \quad \Sigma x_j y_j = 0,$$

which are equivalent to

$$\Sigma z_j^2 = 0, \quad \Sigma z_j \bar{z}_j = 2,$$

where $z_j = x_j + iy_j$. Let $\Pi^{(n-1)}$ be a complex projective space, referred to homogeneous coordinates $z_1, \ldots, z_n$, which are "semi-normalized" by the condition $\Sigma z_j \bar{z}_j = 2$. Let $w(V_{n,2}) = \Pi^{(n-1)}$ be the map which is given by

$$w(x, y) = z,$$

where $z = (z_1, \ldots, z_n)$. Then $w^{-1}(z)$ consists of the circuit in $V_{n,2}$ which is given by

$$x_j + iy_j = e^{i\theta} z_j \quad (0 \leqslant \theta \leqslant 2\pi),$$

and the theorem follows from the second corollary to Theorem 14.

In topological terms, Q is a 2-sphere if $n = 3$, and if $n = 4$ it is the topological product of two 2-spheres, as one sees by referring it to complex parameters, with two generators of opposite systems as axes. If $n > 4$ a complex line on Q constitutes a basis for the (cyclic infinite) homology group $\beta_2(Q)$. Also $\pi_1(Q) = 1$, whence $\pi_2(Q) \simeq \beta_2(Q) = Z_\infty$.

Balliol College,
Oxford.

THE UNIQUENESS OF TRIGONOMETRICAL INTEGRALS

By J. L. B. COOPER.

[Received 10 February, 1941.—Read 29 May, 1942.]

1. *Introduction.* The many investigations which have been made up to now into the conditions under which a trigonometrical integral is unique have dealt with conditions involving convergence at individual points, or summability at individual points. In view of the importance of mean convergence in the theory of trigonometrical integrals, it is of interest to inquire what happens when the integral is only convergent in mean*. This paper is devoted to the investigation of this question.

The main theorem of the paper is the following:

THEOREM I. *If $f(u)$ is integrable over any finite interval, and if, over any finite interval in t,*

$$\lim_{\lambda\to\infty}\int dt\left|\int_{-\lambda}^{\lambda}\left(1-\frac{|u|}{\lambda}\right)^n f(u)\,e^{iut}\,du - F(t)\right| = 0,$$

then, for almost all u,

$$f(u) = \frac{1}{2\pi}\lim_{\lambda\to\infty}\int_{-\lambda}^{\lambda}\left(1-\frac{|t|}{\lambda}\right)^{n+3} F(t)\,e^{-iut}\,dt.$$

This theorem can be deduced from the two following, which are special cases of it.

THEOREM II. *If $f(u)$ is integrable over any finite interval, and if, over any finite interval in t,*

$$\text{l.i.m.}\int_0^{\lambda}\left(1-\frac{u}{\lambda}\right)^n f(u)\cos ut\,du = F(t),$$

* The necessity of theorems of this nature was pointed out in connection with Theorem II in my paper *Proc. London Math. Soc.* (2), 45 (1939), 425–439.

then, for almost all u,

$$f(u) = \frac{2}{\pi} \lim_{\lambda \to \infty} \int_0^\lambda \left(1 - \frac{t}{\lambda}\right)^{n+3} F(t) \cos ut \, dt.$$

THEOREM III. *If $f(u)$ is integrable over any finite interval, and if, over any finite interval in t,*

$$\text{l.i.m.} \int_0^\lambda \left(1 - \frac{u}{\lambda}\right)^n f(u) \sin ut \, du = F(t),$$

then, for almost all u,

$$f(u) = \frac{2}{\pi} \lim_{\lambda \to \infty} \int_0^\lambda \left(1 - \frac{t}{\lambda}\right)^{n+2} F(t) \sin ut \, dt.$$

We add also

THEOREM I*a*. *If the integral $\int_0^\infty f(u)\,du$ is summable (C, n), then the integrals for $f(u)$ in Theorems I, II, III are summable $(C, n+1)$.*

These theorems should be compared with those of Macphail and Titchmarsh*, and those of Offord†. They are a generalisation of the former theorems, since uniform convergence is a special case of mean convergence. They are not a generalisation of Offord's work, which is based on the assumption of summability at every point (or something near to that).

As regards the proofs of the theorems, an important difference from these earlier ones is that while their hypotheses lead to the conclusion, in the cosine case, that $\int_0^\infty f(u)\,du$ exists in a (C, n) sense, ours do not. The first step in the proof is to establish a substitute for this, and we prove in Lemma 3 that, in the cosine case, the integral $\int_0^\infty \frac{f(u)}{u^2}\,du$ exists in a (C, n) sense. This difference explains why our inversion formula is only proved to be summable $(C, n+3)$, whereas the others are summable $(C, n+1)$. The stronger result can be proved in both the sine and in the cosine case if we make the additional assumption that $\int_0^\infty f(u)\,du$ exists in a (C, n) sense.

---

* Macphail and Titchmarsh, *Journal London Math. Soc.*, 11 (1936), 313–318.

† A. C. Offord, *Proc. London Math. Soc.* (2), 42 (1937), 422–480.

Unlike Offord, we do not need the idea of the generalized derivative, because of the convergence in mean of our formula. This is the reason why our theorems are true for $n > 1$, while his are not.

2. *Preliminary lemmas.* The first lemma we prove is an analogue of Lemma 2 of Offord's paper, in which we deal with summability in mean in place of ordinary summability. We shall have occasion to use the original lemma, and shall refer to it as Offord's Lemma 2.

For any integrable function $A(u, t)$ we write

$$A_1(u, t;\ a) = \int_a^u A(u, t)\,du,$$

$$A_{k+1}(u, t) = A_{k+1}(u, t;\ a) = \int_a^u A_k(u, t)\,du.$$

LEMMA 1. *If for any finite α and β, with $\alpha < \beta$,*

$$\int_\alpha^\beta |A_{k+1}(u, t;\ a)|\,du = O(u^{k+m})$$

as $u \to \infty$, and if $b(u)$ is a $k+1$ times differentiable function of u such that

$$b^\nu(u) = O(u^{-\nu-m-1}) \quad (0 \leqslant \nu \leqslant k+1)$$

as $u \to \infty$, then the limit in mean (C, k) over (α, β),

$$\text{l.i.m.}_{\lambda\to\infty} \int_0^\lambda \left(1-\frac{u}{\lambda}\right)^k A(u, t)\,b(u)\,du,$$

exists and is equal to

$$\text{l.i.m.}_{\lambda\to\infty} (-1)^{k+1} \int_0^\lambda A_{k+1}(u, t)\,b^{k+1}(u)\,du.$$

It is shown by Offord that

$$\begin{aligned} I(t, w) &= \int_0^w \left(1-\frac{u}{w}\right)^k A(u, t)\,b(u)\,du \\ &= w^{-k} A_{k+1}(w, t;\ a)\,b(w) \\ &\qquad + \sum_{\nu=1}^{k+1} (-)^\nu \binom{k+1}{\nu} w^{-k+\nu+1} \int_a^w A_{k+1}(w, t)\left(1-\frac{u}{w}\right)^{\nu-1} b^\nu(u)\,du \\ &= I_0(t, w) + \sum_{\nu=1}^{k} I_\nu(t, w) + I_{k+1}(t, w). \end{aligned}$$

Now, for $\nu \leqslant k$,

$$\int_\alpha^\beta |I_\nu(t, w)|\,dt \leqslant Cw^{-k+\nu-1} \int_\alpha^\beta dt \left| \int_a^w A_{k+1}(u, t)\left(1-\frac{u}{w}\right)^{\nu-1} b^\nu(u)\,du \right|$$

$$\leqslant Cw^{-k+\nu-1} \int_a^w u^{-\nu-m-1}\,du \int_\alpha^\beta |A_{k+1}(u, t;\ a)|\,dt$$

$$\leqslant Aw^{-k+\nu-1} \int_a^w u^{-\nu-1+k}\,du$$

$$= O(w^{-1}) \quad \text{if} \quad \nu \leqslant k-1$$

$$\text{or } O(w^{-1}\log w) \quad \text{if} \quad \nu = k.$$

From the hypotheses,

$$\int_\alpha^\beta |I_0(t, w)|\,dt = O(w^{-1}).$$

Hence, in mean,

$$I(t, w) = I_{k+1}(t, w) + O(w^{-1}\log w),$$

so that the limit in mean of $I(t, w)$ exists and is equal to the limit of $I_{k+1}(t, w)$; in view of the restrictions on A and b, $\lim_{w\to\infty} I_{k+1}(t, w)$ exists as an ordinary uniformly convergent integral.

LEMMA 2. *If, as* $\lambda \to \infty$,

$$a_0(\lambda) + a_1(\lambda)\,t + \ldots + a_n(\lambda)\,t^n$$

tends to a limit in mean over a finite interval of values of t, *then*

$$\lim_{\lambda\to\infty} a_\nu(\lambda) = a_\nu$$

exists for all ν, *and the limit in mean is* $\sum_{\nu=0}^{n} a_\nu t^\nu$.

Write

$$a_0(\lambda) + a_1(\lambda)\,t + \ldots + a_n(\lambda)\,t^n = Q(\lambda, t).$$

Choose $n+1$ subintegrals inside the interval of mean convergence, and integrate this equation over each of them. This gives $n+1$ linear equations for the $a(\lambda)$ which can be solved to give their values in terms of the integrals of $Q(\lambda, t)$ over the subintervals, since the discriminant of the equations is not zero. Since the integrals of $Q(\lambda, t)$ tend to limits as $\lambda \to \infty$, the $a_\nu(\lambda)$ also tend to limits. Thus the theorem follows.

3. *Proof of the cosine theorem.* To begin with, let us notice that it is enough to prove the theorem with the assumption $f(u) = 0$ for $u \leqslant 1$; when this special case has been proved, the general theorem is easily deducible from it. For put

$$f(u) = f_1(u) + f_2(u),$$

where $f_1(u) = 0$ $(u \leqslant 1)$, $f_2(u) = 0$ $(u > 1)$. Then $F(u) = F_1(u) + F_2(u)$, and the inversion result follows for $f_1(u)$ because it obeys the special hypothesis, and for $f_2(u)$ because it is $L(0, \infty)$. Hence the result follows for $f(u)$.

For the rest of this section it will be part of the hypotheses of each lemma that $f(u)$ obeys the hypotheses of Theorem II and is zero for $u \leqslant 1$. This is stated here to avoid repetition in each enunciation. We shall write

$$F_1(t) = \int_0^t F(t)\,dt,$$

$$F_k(t) = \int_0^t F_{k-1}(t)\,dt = \frac{1}{(k-1)!}\int_0^t (t-v)^{k-1} F(v)\,dv.$$

LEMMA 3.

$$(a) \qquad \lim_{\lambda\to\infty} (-)^r \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\cos ut}{u^{2r}}\,du = I_{2r}$$

exists for all integral $r \geqslant 1$.

$$(b) \qquad P_{2r}(t) = (-)^r \text{ l.i.m.} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\cos ut}{u^{2r}}\,du,$$

$$P_{2r-1}(t) = (-)^{r-1} \text{ l.i.m.} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\sin ut}{u^{2r-1}}\,du$$

exist for all $r \geqslant 1$; *and for any* t, *and for all* $r \geqslant 1$,

$$\left.\begin{aligned} F_{2r}(t) &= P_{2r}(t) + \sum_{\nu=1}^{r-1} a_{2\nu} t^{2\nu}, \\ F_{2r-1}(t) &= P_{2r-1}(t) + \sum_{\nu=0}^{r-1} a_{2\nu-1} t^{2\nu-1}. \end{aligned}\right\} \qquad (3.1)$$

From the hypotheses, for all t,

$$\lim_{\lambda\to\infty} \int_0^t (t-v)^\nu\,dv \left\{F(v) - \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n f(u)\cos uv\,du\right\} = 0.$$

Hence, for all t,

$$F_{2r}(t) = (-1)^r \lim_{\lambda\to\infty} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n f(u)\, u^{-2r} \left\{\cos ut - 1 + \ldots + (-)^r \frac{u^{2r-2}\, t^{2r-2}}{(2r-2)!}\right\} du$$

$$= (-1)^r \lim_{\lambda\to\infty} \left\{\int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\cos ut}{u^{2r}}\, du \right.$$

$$\left. + \sum_{\nu=0}^{r-1} (-)^{\nu-1} \frac{t^{2\nu}}{2\nu!} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)}{u^{2(r-\nu)}}\, du\right\} \qquad (3.2)$$

if $\nu = 2r-1$, and

$$F_{2r-1}(t) = (-1)^{r-1} \lim_{\lambda\to\infty} \left\{\int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\sin ut}{u^{2r-1}}\, du \right.$$

$$\left. - \sum_{\nu=0}^{r-1} (-)^\nu \frac{t^{2\nu-1}}{(2\nu-1)!} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)}{u^{2(r-\nu)}}\, du\right\} \qquad (3.3)$$

if $\nu = 2r-2$.

If in Lemma 1 we take $A(u,t) = f(u)\cos ut$, or $f(u)\sin ut$, $b(u) = u^{-\nu}$ $(u \geqslant 1)$, $k = n$, $m = 0$, it follows that $P_\nu(t)$ exists as a limit in mean. The right-hand sides of (3.2) and (3.3) are boundedly convergent, hence convergent in mean, over any finite interval in t; and so the limits in mean of the polynomial sums exist. By Lemma 2, the limit of each separate term exists. This gives part (a) of the lemma; part (b) then follows from (3.2) and (3.3).

Write

$$\phi_1(u) = \int_1^u \frac{f(v)}{v^2}\, dv, \quad \phi_{k+1}(u) = \int_1^u \phi_k(v)\, dv.$$

From the lemma above, we know that $\phi_{n+1}(u)$ is $O(u^n)$.

LEMMA 4. (a) *If* $n = 2m$, $2r \geqslant 2m+4$,

$$P_{2r}(t) = (-)^{n+1+r} \int_1^\infty \phi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \left(\frac{\cos ut}{u^{2r-2}}\right) du; \qquad (3.4)$$

if $n = 2m+1$, $2r \geqslant n+4$,

$$P_{2r-1}(t) = (-)^{n+r} \int_1^\infty \phi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \left(\frac{\sin ut}{u^{2r-3}}\right) du. \qquad (3.5)$$

(b) *If* $r \geqslant n+4$, $P_r(t) = o(t^{n+1})$.

We have

$$P_{2r}(t) = (-)^r \text{ l.i.m.} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\cos ut}{u^{2r}}\,du, \tag{3.6}$$

In Offord's Lemma 2, take $k=n$, $m=0$,

$$A(u) = f(u)/u^2, \quad b(u) = \cos ut/u^{2r-3};$$

then the conditions are satisfied if $2r \geqslant n+4$, since $\phi_{n+1}(u)$ is $O(u^n)$. The lemma then shows that the limit of the integral on the right of (3.6) is equal to the right hand of (3.4). Since the limit of a function is equal almost everywhere to its limit in mean, (3.4) follows. (3.5) follows similarly by taking

$$A(u) = f(u)/u^2, \quad b(u) = \sin ut/u^{2r-3},$$

provided that $2r-1 \geqslant n+4$.

From (3.4)

$$P_{2r}(t) = \pm t^{n+1} \int_1^\infty \frac{\phi_{n+1}(u)}{u^{n+2}} \cos ut\,du + O(t^n):$$

so $P_{2r}(t)$ is $o(t^{n+1})$, by the Riemann-Lebesgue theorem. Similarly, $P_{2r-1}(t)$ is $o(t^{n+1})$.

LEMMA 5. *(a) If $n=2m$, then almost everywhere*

$$(-)^m \frac{f(x)}{x^{2m+4}} = \lim_{\mu\to\infty} \frac{2}{\pi} \int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} P_{2m+4}(t)\cos tx\,dt. \tag{3.7}$$

(b) If $n=2m+1$, then almost everywhere

$$(-)^m \frac{f(x)}{x^{2m+5}} = \lim_{\mu\to\infty} \frac{2}{\pi} \int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} P_{2m+5}(t)\sin tx\,dt. \tag{3.8}$$

We have, from Lemma 4,

$$\int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} P_{2m+4}(t)\cos tx\,dt$$

$$= (-)^{n+1} \int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} \cos tx\,dt \int_1^\infty \phi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \left(\frac{\cos ut}{u^{2m+2}}\right) du$$

$$= (-)^{n+1} \int_1^\infty \phi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \frac{1}{u^{2m+2}} \int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} \cos tx\cos ut\,du.$$

Now

$$\int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} \cos tx\cos ut\,du = \tfrac{1}{2}\{K_{n+1}(u+x,\mu)+K_{n-1}(u-x,\mu)\},$$

where

$$K_{n+1}(u,\mu) = \mu^{-n-1}\int_0^\mu (\mu-t)^{n+1}\cos ut\,du$$

$$= (n+1)!\,\mu^{-n-1}u^{-n-2}I^{n+2}(\mu u),$$

and $I^{n+2}(v)$ is the integral of order $n+2$ of $\cos v$. Also

$$\left(\frac{d}{du}\right)^{n+1}\left\{\frac{K_{n+1}(u-x,\mu)}{u^{2m+2}}\right\} = \left(\frac{d}{du}\right)^{n+1}\left\{\frac{(n+1)!\,I^{n+1}\{\mu(u-x)\}}{u^{2m+2}\mu^{n+1}(u-x)^{n+2}}\right\}$$

$$= (n+1)!\,\frac{\sin\mu(u-x)}{u^{2m+2}(u-x)^{n+2}} + O\left(\frac{1}{\mu u^{2m+n+4}}\right).$$

It is easy to see that

$$\left\{\int_1^{x-\delta}+\int_{x+\delta}^{\infty}\right\}\phi_{n+1}(u)\left(\frac{d}{du}\right)^{n+1}\frac{1}{u^{2m+2}}\{K_{n+1}(u+x,\mu)+K_{n+1}(u-x,\mu)\}\,du$$

$$+\int_{x-\delta}^{x+\delta}\phi_{n+1}(u)\left(\frac{d}{du}\right)^{n+1}\frac{K_{n+1}(u+x,\mu)}{u^{2m+2}}\,du$$

tends to zero as $\mu\to\infty$, in consequence of the Riemann-Lebesgue theorem. It remains to consider

$$\int_{x-\delta}^{x+\delta}\phi_{n+1}(u)\left(\frac{d}{du}\right)^{n+1}\frac{K_{n+1}(u-x,\mu)}{u^{2m+2}}\,du$$

$$= \left[\sum_{\nu=0}^{n}(-)^{n-\nu+1}\phi_{\nu+1}(u)\left(\frac{d}{du}\right)^{\nu}\frac{K_{n+1}(u-x,\mu)}{u^{2m+2}}\right]_{x-\delta}^{x+\delta}$$

$$+(-)^{n+1}\int_{x-\delta}^{x+\delta}\frac{f(u)}{u^{2m+4}}K_{n+1}(u-x,\mu)\,du.$$

The expression in brackets is $O(\mu^{-1})$ as $\mu\to\infty$, and hence tends to zero. The integral is

$$\int_{x-\delta}^{x+\delta}\frac{f(u)}{u^{2m+4}}\,du\int_0^\mu\left(1-\frac{t}{\mu}\right)^{n+1}\cos t(u-x)\,dt \to \pi\,\frac{f(x)}{x^{2m+4}}$$

almost everywhere, by Fejér's Theorem. Hence the original integral tends to $(-1)^m f(x)/x^{2m+4}$, which gives part (a) of the lemma. The proof of part (b) is along precisely the same lines.

LEMMA 6. *For almost all* $x \geqslant 1$,

$$f(x) = \lim_{\lambda\to\infty}\frac{2}{\pi}\int_0^\mu\left(1-\frac{t}{\mu}\right)^{n+3}\cos tx\,F(t)\,dt. \qquad (3.9)$$

Consider

$$\int_0^\mu (\mu-t)^{n+3} F(t) \cos tx\, dt$$

$$= \left[\sum_{\nu=0}^{n+4} (-)^{\nu-1} P_\nu(t) \left(\frac{d}{dt}\right)^{\nu-1} \{(\mu-t)^{n+3} \cos tx\}\right]_0^\mu$$

$$+ (-)^{n+4} \int_0^\mu P_{n+4}(t) \left(\frac{d}{dt}\right)^{n+4} \{(\mu-t)^{n+3} \cos tx\}\, dt. \quad (3.10)$$

In the bracket, the contribution of the terms at $t = 0$ is $O(\mu^{n+2})$, because $P_\nu(0)$ is zero for odd ν. The contribution of $t = \mu$ is

$$(-)^{n+3} P_{n+4}(\mu)(n+3)! \cos \mu x;$$

and this is $o(\mu^{n+1})$, by Lemma 4 (b). Since to get (3.9) we multiply (3.10) by μ^{-n-3}, these terms give no contribution to (3.9). The integral in (3.10) is, if $n = 2m$,

$$(-)^m x^{2m+4} \int_0^\mu P_{2m+4}(t)(\mu-t)^{n+3} \cos tx\, dt \quad (3.11)$$

plus integrals of the form

$$x^P \int_0^\mu P_{2m+4}(\mu-t)^\nu \begin{matrix}\cos\\ \sin\end{matrix} tx\, dt, \quad (3.12)$$

where $\nu < n+3$; and if $n = 2m+1$, the integral is

$$(-)^m x^{2m+5} \int_0^\mu P_{2m+5}(t)(\mu-t)^{n+3} \cos tx\, dt \quad (3.13)$$

plus remainder terms of the form (3.12). From Lemma 5, we see that all that is necessary to complete the proof is to show that all the terms (3.12) are $o(\mu^{n+3})$, for on division by μ^{n+3} the integrals (3.11) and (3.13) tend to the right-hand side of (3.9). The integrals (3.12) are of the form

$$\pm \int_1^\infty \phi_{n+1}(u)\, du \left(\frac{d}{du}\right)^{n+1} \frac{1}{u^{n+2}} \int_0^\mu (\mu-t)^\nu \begin{matrix}\cos\\ \sin\end{matrix} ut \begin{matrix}\cos\\ \sin\end{matrix} tx\, dt. \quad (3.14)$$

If in the inner integral we take the sines or the cosines together, it is

$$\tfrac{1}{2}\mu^\nu \{K_\nu(u-x, \mu) \pm K_\nu(u+x, \mu)\}, \quad (3.15)$$

and if we take a sine with a cosine, it is

$$\tfrac{1}{2}\mu^\nu \{K_\nu'(u-x, \mu) \pm K_\nu'(u+x, \mu)\},$$

where

$$\mu^\nu K'(u,\mu) = \int_0^\mu (\mu - t)^\nu \sin ut\, dt = \frac{\mu^\nu}{u} - \frac{\nu}{u}\mu^{\nu-1} K_{\nu-1}(u,\mu),$$

on integrating by parts. Since terms of order μ^ν contribute nothing in the limit, the whole problem reduces to considering kernels of the form (3.15).

With this understanding, break up the integral (3.14) into the sum $\int_1^{x-\delta} + \int_{x-\delta}^{x+\delta} + \int_{x+\delta}^{\infty}$. Then

$$\int_{x+\delta}^{\infty} = \mu^{n+1} \int_{x+\delta}^{\infty} \frac{\phi_{n+1}(u)}{u^{n+2}} \left\{ \frac{\cos \mu(u+x)}{(u+x)^{\nu+1}} \pm \frac{\cos \mu(u-x)}{(u-x)^{\nu-1}} \right\} du + O(\mu^n) = o(\mu^{n+1}).$$

Similarly,

$$\int_1^{x-\delta} = o(\mu^{n+1}),$$

and

$$\int_{x-\delta}^{x+\delta} \phi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \left\{ \frac{1}{u^{n+2}} K_\nu(u+x, \mu) \right\} du = o(\mu^{n+1}).$$

There remain only the terms

$$\mu^\nu \int_{x-\delta}^{x+\delta} \phi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \frac{K_\nu(\mu, u-x)}{u^{n+2}}\, du$$

$$= \mu^\nu \left[\sum_{k=0}^{n+1} (-)^{n+1-k} \phi_k(u) \left(\frac{d}{du}\right)^{k+1} \frac{K_\nu(\mu, u-x)}{u^{n+2}} \right]_{x-\delta}^{x+\delta}$$

$$+ \mu^\nu \int_{x-\delta}^{x+\delta} \frac{f(u)}{u^{n+3}} K_\nu(\mu, u-x)\, du.$$

The bracket is $O(\mu^n)$, and for almost all x the integral is $O(\mu^\nu)$; hence both are $o(\mu^{n+3})$.

This completes the proof of the lemma, and so the cosine theorem is proved.

4. *Proof of the sine theorem.* The proof of the sine theorem is so similar to that of the last section that it seems unnecessary to give the reasoning in full; it will be sufficiently indicated by giving the enunciations of the lemmas corresponding to those above.

In what follows $f(u)$ is supposed to satisfy the hypotheses of Theorem III and to be zero in (0, 1).

LEMMA 3′.

$$(a) \qquad \lim_{\lambda \to \infty} \int_1^\lambda \left(1 - \frac{u}{\lambda}\right)^n \frac{f(u)}{u^{2\nu-1}}\, du = I_{2\nu-1}$$

exists for all integral $\nu \geqslant 1$;

$$(b) \qquad Q_{2r}(t) = (-)^r \text{ l.i.m.} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\sin ut}{u^{2r}}\,du,$$

$$Q_{2r-1}(t) = (-)^r \text{ l.i.m.} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n \frac{f(u)\sin ut}{u^{2r-1}}\,du$$

exist for all $r \geqslant 1$, *and*

$$F_{2r}(t) = Q_{2r}(t) + \sum_{\nu=1}^{r} b_{2\nu-1} t^{2\nu-1}, \quad F_{2r-1}(t) = Q_{2r-1}(t) + \sum_{\nu=0}^{r-1} b_{2\nu} t^{2\nu}.$$

We write

$$\psi_1(u) = \int_1^u \frac{f(v)}{v}\,dv, \quad \psi_{k+1}(u) = \int_1^u \psi_k(v)\,dv,$$

Then, from Lemma 3′ (*a*), $\psi_{n+1}(u) = O(u^n)$.

LEMMA 4′. (*a*) *If* $n = 2m$, $2r > 2m+4$,

$$Q_{2r-1}(t) = (-)^{r+1} \int_1^\infty \psi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \left(\frac{\cos ut}{u^{2r-2}}\right) du;$$

if $n = 2m+1$, $2r \geqslant 2m+4$,

$$Q_{2r}(t) = (-)^r \int_1^\infty \psi_{n+1}(u) \left(\frac{d}{du}\right)^{n+1} \left(\frac{\sin ut}{u^{2r-1}}\right) du.$$

(*b*) *If* $r \geqslant n+3$, $P_r(t) = o(t^{n+1})$.

LEMMA 5′. *For almost all* $x > 1$,

(*a*) *if* $n = 2m$,

$$(-)^m \frac{f(x)}{x^{2m+3}} = \lim_{\mu\to\infty} \frac{2}{\pi} \int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} P_{n+3}(t)\cos tx\,dt;$$

(*b*) *if* $n = 2m+1$,

$$(-)^m \frac{f(x)}{x^{2m+3}} = \lim_{\mu\to\infty} \frac{2}{\pi} \int_0^\mu \left(1-\frac{t}{\mu}\right)^{n+1} P_{n+3}(t)\sin tx\,dt.$$

LEMMA 6′. *For almost all* $x > 1$,

$$f(x) = \lim_{\mu\to\infty} \frac{2}{\pi} \int_0^\infty \left(1-\frac{t}{\mu}\right)^{n+2} \sin tx\, F(t)\,dt.$$

In the proof of this last lemma, the fact that we know that $Q_{n+3}(t)$ is $o(t^{n+1})$ enables us to use only $n+3$ integrations by parts in the formula corresponding to (3.10), whereas we need $n+4$ in the cosine case, because

this can be proved there only for $P_{n+4}(t)$. This is the reason why we can prove the inversion formula with summability $(C, n+2)$ in this case, and only with summability $(C, n+3)$ in the cosine case. This point will be discussed further later on.

5. *The proof of the main theorem.* It must now be shown that Theorem I follows from Theorems II and III. On the hypotheses of Theorem I,

$$\text{l.i.m.} \int_{-\lambda}^{\lambda} \left(1-\frac{|u|}{\lambda}\right)^n f(u)\, e^{iut}\, du = F(t),$$

and so

$$\text{l.i.m.} \int_0^{\lambda} \left(1-\frac{u}{\lambda}\right)^n [f(u)\, e^{iut}+f(-u)\, e^{-iut}]\, du = F(t),$$

$$\text{l.i.m.} \int_0^{\lambda} \left(1-\frac{u}{\lambda}\right)^n [f(u)\, e^{-iut}+f(-u)\, e^{iut}]\, du = F(-t).$$

On adding and subtracting these last formulae, we get

$$\text{l.i.m.} \int_0^{\lambda} \left(1-\frac{u}{\lambda}\right)^n [f(u)+f(-u)] \cos ut\, du = \tfrac{1}{2}\{F(t)+F(-t)\},$$

$$\text{l.i.m.} \int_0^{\lambda} \left(1-\frac{u}{\lambda}\right)^n [f(u)-f(-u)] \sin ut\, du = \frac{1}{2i}\{F(t)-F(-t)\}.$$

Thus, for almost all u, by Theorems 2 and 3,

$$f(u)+f(-u) = \frac{1}{\pi}\int_0^{\infty} [F(t)+F(-t)] \cos ut\, dt \quad (C, n+3),$$

$$f(u)-f(-u) = -\frac{i}{\pi}\int_0^{\infty} [F(t)-F(-t)] \sin ut\, dt \quad (C, n+3),$$

and so

$$f(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(t)\, e^{-iut}\, dt \quad (C, n+3).$$

A question raised by comparing Theorem III with the theorems of Macphail and Titchmarsh and of Offord is whether the order of summability of the inversion formula can be improved; for in the latter theorem the order is $(C, n+1)$, and in ours $(C, n+3)$.

Consider the proof of the cosine theorem. Suppose that we tried to prove Lemma 6 with $(\mu-t)^{n+1}$ in place of $(\mu-t)^{n+3}$. On the integration by parts in (3.10), we would get from the bracket at $t=\mu$ the terms

$$P_{n+2}(\mu) \cos \mu x + x P_{n+3}(\mu) \sin \mu x.$$

The other terms in the bracket and the integrals would be $o(\mu^{n+1})$, apart from the integrals which give the result we want after division by μ^{n+3}.

The question whether the integral is summable $(C, n+1)$ depends on whether $P_{n+2}(\mu)$ and $P_{n+3}(\mu)$ are each $o(\mu^{n+1})$. In the same way, summability $(C, n+2)$ depends on whether $P_{n+3}(\mu)$ is $o(\mu^{n+2})$.

To take, for example, the case n even, to prove summability $(C, n+1)$ we must prove that, if

$$\text{l.i.m.} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n f(u) \cos ut\, du$$

exists over every finite interval in t, then

$$\lim_{\lambda\to\infty} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n f(u) \frac{\cos ut}{u^{n+2}}\, du$$

and

$$\lim_{\lambda\to\infty} \int_1^\lambda \left(1-\frac{u}{\lambda}\right)^n f(u) \frac{\sin ut}{u^{n+3}}\, du$$

are $o(t^{n+1})$ as $t\to\infty$. So far I have not been able either to prove this, or to find an example to show it to be false.

If, however, we make the additional assumption that the integral $\int_0^\infty f(u)\, du$ is summable (C, n), it is easy to see that the method of Lemma 4 is then applicable to $P_{n+2}(t)$ and to $P_{n+3}(t)$, and the required order results follow for these functions. In this case summability $(C, n+1)$ can be proved: the theorems of Macphail and Titchmarsh, in the general case, are particular cases of this result, which is Theorem I*a*.

The part played by the modulus sign in the theorem—that is, the rôle of the assumption of strong convergence—appears if we consider what happens if the integral for $F(t)$ is only supposed to converge weakly. We take the case $F(t)$ zero, *i.e.* the uniqueness theorem.

In this case we assume only that, for all finite β and y,

$$\lim_{\lambda\to\infty} \int_y^{\beta+y} dt \left\{ \int_{-\lambda}^{\lambda} \left(1-\frac{|u|}{\lambda}\right)^\alpha f(u)\, e^{iut}\, du \right\} = 0.$$

From this we deduce that

$$\lim_{\lambda\to\infty} \int_{-\lambda}^{\lambda} \left(1-\frac{|u|}{\lambda}\right)^\alpha f(u)\, e^{iuy} \frac{e^{i\beta u}-1}{iu}\, du = 0.$$

If $\alpha \leqslant 1$, Offord's uniqueness theorem gives

$$f(u) \frac{e^{i\beta u}-1}{iu} = 0,$$

for almost all u and all $\beta > 0$. Hence $f(u) = 0$ almost everywhere. However, if $\alpha > 1$, uniqueness breaks down, as we see by taking $f(u) = u^2$. Thus the modulus sign is essential for uniqueness to hold for $\alpha > 1$.

6. *Uniqueness with general kernels.* Uniqueness theorems similar to those of the last sections can be proved for summation by means of a class of kernels more general than those used in Cesàro summation.

We denote by $k(\lambda, u)$ a kernel which satisfies the following conditions:

(i) *for some M, and for all w and u,*

$$|k(w, u)| \leqslant M; \tag{6.1}$$

(ii) *for all u, $\lim k(w, u) = 1$ as $w \to \infty$.*

The class of kernels k plainly includes the Cesàro kernels $(1-|u|/\lambda)^\alpha$, Abel's kernel $e^{-u/\lambda}$, Weierstrass's kernel $e^{-u^2/\lambda}$, and so on.

In order to prove the uniqueness theorems, with the means at our disposal, for a general kernel k, it is necessary to restrict either the size of $f(u)$ at infinity or the manner in which the trigonometrical integral is summable in mean. In this section two theorems, each with one of the alternative restrictions, will be proved.

In the proofs of the theorems a lemma, similar to Lemma 2, will be useful.

LEMMA 7. *If a sequence of polynomials of the same degree n tends to a limit for almost all values of its argument, the sequences consisting of the coefficients of a given power all tend to limits, and the limit of the sequence of polynomials is a polynomial of degree n.*

This is a consequence of well-known theorems. It can be proved by an obvious modification of the proof of Lemma 2.

THEOREM IV. *If*

(i) $\dfrac{f(u)}{1+|u|^n}$ *is* $L(-\infty, \infty)$ *for some* n;

(ii) $k(\lambda, u)f(u)$ *is* $L(-\infty, \infty)$ *for all* λ;

(iii) *for every finite interval in t,*

$$\lim_{\lambda\to\infty} \int dt \left| \int_{-\infty}^{\infty} k(\lambda, u) f(u) e^{iut}\, du \right| = 0; \tag{6.2}$$

then $f(u)$ is zero almost everywhere.

Since $k(\lambda, u)f(u)$ is $L(-\infty, \infty)$, the integral $\int_{-\infty}^{\infty} k(\lambda, u) f(u) e^{iut}\, du$ can be integrated under the sign of integration. On integrating it n times

over the interval $(0, t)$, we get, from (6.2),

$$\lim_{\lambda\to\infty}\int_{-\infty}^{\infty} k(\lambda, u)\, f(u)\, \frac{1}{(iu)^n}\left\{e^{iut}-1-iut-\ldots-\frac{(iut)^{n-1}}{(n-1)!}\right\} du = 0.$$

Hence, for any $a > 0$,

$$\lim_{\lambda\to\infty}\left\{\left(\int_{-\infty}^{\infty}+\int_{a}^{\infty}\right) k(\lambda, u)\, \frac{f(u)}{(iu)^n}\, e^{iut}\, du\right.$$

$$+\int_{-a}^{a} k(\lambda, u)\, f(u)\, \frac{1}{(iu)^n}\left\{e^{iut}-1-iut-\ldots-\frac{(iut)^{n-1}}{(n-1)!}\right\} du$$

$$\left.+\sum_{\nu=0}^{n-1}\frac{t^\nu}{\nu!}\left(\int_{-\infty}^{-a}+\int_{a}^{\infty}\right) k(\lambda, u)\, f(u)\, \frac{du}{(iu)^{n-\nu}}\right\} = 0. \qquad (6.3)$$

The limit of the first integral is

$$F_1(t) = \left(\int_{-\infty}^{-a}+\int_{a}^{\infty}\right)\frac{f(u)}{(iu)^n}\, e^{iut}\, du,$$

since $(iu)^{-n} k(\lambda, u) f(u)$ tends to $(iu)^{-n} f(u)$ and is dominated by $M|u|^{-n}|f(u)|$, which is integrable over the ranges $(-\infty, -a)$ and (a, ∞). Similarly, the limit of the second integral is

$$F_2(t) = \int_{-a}^{a} f(u)\, \frac{1}{(iu)^n}\left\{e^{iut}-1-\ldots-\frac{(iut)^{n-1}}{(n-1)!}\right\} du.$$

Thus the limit of the sum in (6.3) exists for all t. Since this sum is a polynomial of degree $n-1$ in t, the limit is also a polynomial of degree $n-1$. We have therefore

$$F_1(t)+F_2(t)+\sum_{\nu=0}^{n-1} c_\nu t^\nu = 0. \qquad (6.4)$$

Now, for almost all $x > a$,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} F_1(t)\, e^{-ixt}\, dt = \frac{f(x)}{(ix)^n} \quad (C, 1), \qquad (6.5)$$

by the Fejér-Lebesgue theorem, since $F_1(t)$ is the transform of a function of $L(-\infty, \infty)$. Consider now

$$\lim_{\mu\to\infty}\int_{-\mu}^{\mu}\left(1-\frac{|t|}{\mu}\right)^{n+1} F_2(t)\, e^{-ixt}\, dt.$$

The integral is

$$\int_{-a}^{a}\frac{f_2(u)}{(iu)^n}\, du\int_{-\mu}^{\mu}\left(1-\frac{|t|}{\mu}\right)^{n+1} e^{-ixt}\left(e^{iut}-1-iu-\ldots-\frac{(iut)^{n-1}}{(n-1)!}\right) du. \qquad (6.6)$$

Write

$$2K_{n+1}(\mu, x) = \int_{-\mu}^{\mu} \left(1-\frac{|t|}{\mu}\right)^{n+1} e^{-ixt}\,dt$$

$$= 2\int_0^{\mu} \left(1-\frac{t}{\mu}\right)^{n+1} \cos xt\,dt = 2\mu^{-n-1}x^{-n-2}I^{n+2}(\mu x),$$

where $I^{n+2}(u)$ is the integral of order $n+2$ of $\cos u$. Then the inner integral in (6.6) is

$$2\left\{K_{n+1}(x-u) - uK'_{n+1}(x) - \frac{u^2}{2!}K''_{n+1}(x) - \ldots - \frac{u^{n-1}}{(n-1)!}K^{n-1}_{n+1}(x)\right\}$$

$$= 2\frac{u^n}{n!}K^n_{n+1}(x-\theta u),$$

by Taylor's Theorem. Here we have written $K^{\nu}_{n+1}(x)$ for

$$\left(\frac{d}{dx}\right)^{\nu} K_{n+1}(\mu, x).$$

For all $x > a$,

$$|K^n_{n+1}(x-\theta u)| < A\left|\frac{\sin\mu(x-\theta u)}{\mu(x-\theta u)^{n+2}}\right| + O\left(\frac{1}{\mu^2}\right) < \frac{A}{\mu},$$

so that the integral in (6.6) is less than

$$\frac{A}{\mu}\int_{-a}^{a} |f(u)|\,du, \tag{6.7}$$

which tends to zero as $\mu \to \infty$.

Finally, with regard to the sum in (6.4), we know that

$$\int_{-\infty}^{\infty} t^{\nu} e^{ixt}\,dt = 0 \quad (C, n+1) \tag{6.8}$$

for all $\nu \leqslant n$, $x \neq 0$. Together, (6.4), (6.5), (6.7) and (6.8) give

$$x^{-n}f(x) = 0$$

for almost all $x > a$. Since a is arbitrary, this proves that $f(x)$ is zero almost everywhere.

THEOREM V. *If*

(i) $k(\lambda, u)f(u)$ *is* $L(-\infty, \infty)$ *for all* λ;

(ii) *for some* $n > 0$,

$$\lim_{\lambda\to\infty}\int_{-\infty}^{\infty}\frac{dt}{1+|t|^n}\left|\int_{-\infty}^{\infty} k(\lambda, u)\,f(u)\,e^{iut}\,du\right| = 0;$$

then $f(u) = 0$ *almost everywhere.*

Write $$f_\lambda(u) = k(\lambda, u) f(u)$$

and $$P_\lambda(t) = \int_{-\infty}^{\infty} f_\lambda(u)\, e^{iut}\, dt.$$

Since $f_\lambda(u)$ is $L(-\infty, \infty)$ we can multiply the Fourier integral for $f_\lambda(u)$ by a function of bounded variation and integrate under the sign of integration over any finite interval. Thus

$$\int_0^u (u-v)^{n-1} f_\lambda(v)\, dv = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_\lambda(t)\, dt \int_0^u (u-v)^{n-1} e^{-ivt}\, dv$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} P_\lambda(t)(-it)^{-n} \left\{ e^{-iut} - \sum_{m=0}^{n-1} \frac{(-iut)^m}{m!} \right\} dt.$$

The right-hand side, considered as the limit of an integral over a finite range of t, is the limit of a sequence of polynomials in u; using Lemma 7, we may write it as

$$\int_0^u (u-v)^{n-1} f_\lambda(v)\, dv = \frac{1}{2\pi} \int_{-1}^{1} P_\lambda(t)(-it)^{-n} \left\{ e^{-iut} - \sum_{m=0}^{n-1} \frac{(-iut)^m}{m!} \right\} dt$$

$$+ \frac{1}{2\pi} \left\{ \int_{-\infty}^{-1} + \int_1^{\infty} \right\} P_\lambda(t)\, \frac{e^{-iut} - 1}{(-it)^n}\, dt$$

$$- \frac{i^n}{2\pi} \sum_{m=1}^{n-1} \frac{(-iu)^m}{m!} \left\{ \int_{-\infty}^{-1} + \int_1^{\infty} \right\} P_\lambda(t)\, t^{m-n}\, dt$$

$$= I_1(\lambda, u) + I_2(\lambda, u) + \sum_{m=1}^{n-1} u^m J_m(\lambda).$$

Because of hypothesis (ii), both $I_1(\lambda, u)$ and $I_2(\lambda, u)$ tend to zero, as $\lambda \to \infty$, for all u. The left-hand side tends to

$$\int_0^u (u-v)^{n-1} f(v)\, dv,$$

since $f_\lambda(v)$ tends to $f(v)$ and is dominated by $M|f(v)|$. Hence, as $\lambda \to \infty$,

$$\sum_{m=1}^{n-1} u^m J_m(\lambda) \to \int_0^u (u-v)^{n-1} f(v)\, dv. \qquad (6.9)$$

The left-hand side is a polynomial of degree $n-1$, and, since it tends to a limit everywhere, this limit is also a polynomial of degree $n-1$. The right-hand side is an integral of order n, and hence is $o(t^{n-1})$, as t tends

to zero, so that the limit of the left-hand side must also be $o(t^{n-1})$; but, since the left-hand side is a polynomial of degree $n-1$, this is possible only if it is identically aero. It follows that the right-hand side of (6·9) which is the integral of order n of $f(v)$, is identically zero; hence $f(v)$ is zero almost everywhere.

In conclusion, I wish to express my thanks to Prof. Titchmarsh for his assistance in the preparation of this paper. Also I wish to thank The Queen's College, Oxford, for a Research Grant during the tenure of which this work was done.

3a Elton Road,
Bristol, 8.

PROJECTIVE CONVERGENCE AND LIMIT IN SEQUENCE SPACES

By H. S. ALLEN.

[Received 13 November, 1940.—Read 17 December, 1940.]

Introduction.

It is proposed to study convergence and limit in the linear coordinate spaces of infinite dimensions as distinct from the abstract spaces in which there are no coordinates.

The sequence $x = (x_k)$ is considered either as a point or as a vector issuing from the origin. If x, y are sequences and r is any number, we put $x+y = (x_k+y_k)$, $rx = (rx_k)$ and $xy = \Sigma x_k y_k$, whenever this series is absolutely convergent.

A set S of sequences containing the origin and such that for every x and y in S and every number r, $x+y$ and rx are in S is called a sequence space. The space of all sequences is denoted by σ; some of the subspaces of σ are:

(1) ϕ, the space of finite sequences; (K.T., 196)†

(2) σ_r $(r \geqslant 1)$, in which $\sum_k |x_k|^r$ is convergent; (K.T., 196)

(3) σ_∞, the space of bounded sequences; (K.T., 218)

(4) Γ, the space of convergent sequences; (L.)

(5) Z, the space of nul sequences; (L.)

(6) C, the space of stationary sequences in which $x_{k+1} = x_k$ for $k \geqslant l$; (L.)

(7) D, the space of alternating stationary sequences in which $x_{k+1} = -x_k$ for $k \geqslant l$.

† References in this form are to pages in the paper by Köthe and Toeplitz mentioned in the bibliography. The subsequent abbreviation "(L.)" refers to lectures given by Dr. P. Dienes at Birkbeck College, London.

Definitions.

(1) A set X in σ_∞ is *collectively bounded* when $|x_k| \leqslant M$ for every k and x in X.

(2) A set X in σ_1 is *collectively absolutely convergent* when to every $\epsilon > 0$ corresponds an $m(\epsilon)$ such that $\sum_{k=m+1}^{\infty} |x_k| \leqslant \epsilon$ for every x in X.

(3) A set X in Γ is *collectively convergent* when to every $\epsilon > 0$ corresponds an $N(\epsilon)$ such that, for every x in X, $|x_p - x_q| \leqslant \epsilon$ for $p, q \geqslant N$.

(4) In ϕ, the index of the last non-zero coordinate of x is called the *length*. A set X is of *bounded length* when $x_k = 0$ for $k > l$ for every x in X.

(5) A sequence space α is *normal* when, if x is in α and $|y_k| \leqslant |x_k|$ for every k, then y is in α. (K.T., 198.)

Evidently ϕ, σ_r, σ_∞ and Z are normal; Γ, C, D are not normal.

(6) The set of points y for which $\sum_k |x_k y_k|$ is convergent for every x in a space α forms a linear space α^*, called the *dual space of* α. (K.T., 197.) We have

(i) $\phi^* = \sigma$ and $\sigma^* = \phi$; (L.)

(ii) $\sigma_\infty{}^* = \sigma_1$ and $\sigma_1{}^* = \sigma_\infty$; (K.T., 218)

(iii) If $r > 1$, $\sigma_r{}^* = \sigma_s$ where $\frac{1}{r} + \frac{1}{s} = 1$; (K.T., 219)

(iv) $\Gamma^* = \sigma_1$, $Z^* = \sigma_1$, $C^* = \sigma_1$, $D^* = \sigma_1$. (L.)

(7) The space α is *complete* when $\alpha^{**} = \alpha$. (K.T., 197.) ϕ, σ_r $(r \geqslant 1)$, σ_∞ and σ are complete.

It is known that every complete space is normal and that Z is normal but not complete. Every complete space contains ϕ. α^* is complete and α^{**} is the smallest complete space containing α. If $\alpha \geqslant \beta$, then $\alpha^* \leqslant \beta^*$. (K.T., 197–198.)

We consider convergence in the sequence spaces and first we have, from K.T., 198,

(8) A sequence of sequences $x^{(n)}$ is *coordinate convergent* (c-cgt) when $\lim_{n\to\infty} x_k^{(n)} = x_k$ exists for every k; and then $x = (x_k)$ is the *coordinate limit* (c-limit) of $x^{(n)}$.

(9) In a coordinate space of finite dimensions, if y is at unit distance from the origin, the projection of x on the direction of y is $\Sigma x_k y_k$. If $x^{(n)}$ is a sequence of points in such a space, the sequence of projections of $x^{(n)}$ on any assigned direction is convergent if, and only if, the points $x^{(n)}$ are c-cgt. In sequence spaces of infinite dimensions the coordinate x_k may be considered as the projection of x upon the corresponding axis, and *the projection of x on y is defined as* $\Sigma x_k y_k$, wherever this is absolutely convergent. Thus the set of points which can be projected on every direction in α is the dual space α^*.

(10) If β is a subspace of α^* containing ϕ and if the projections of $x^{(n)}$ in α on every fixed direction in β are convergent, *i.e.* if the sequence $u_n' = \sum_k u_k x_k^{(n)}$ is convergent for every u in β, we say that $x^{(n)}$ is *projective convergent relative to* β, or $\alpha\beta$-convergent, and when $\beta = \alpha^*$ that $x^{(n)}$ is projective convergent in α. This definition coincides with the definition of a convergent sequence of points given by K.T., 197 and 200. We postulate that β shall contain ϕ to ensure the coordinate convergence of $x^{(n)}$. The point in which the k-th coordinate is 1 and the rest are zero is denoted throughout by e_k. The projection of $x^{(n)}$ on e_k is $x_k^{(n)}$, and so projective convergence implies coordinate convergence.

(11) When the projections of a set X in α on every u in β are bounded, X is said to be *projective bounded relative to* β, or $\alpha\beta$-bounded, and so every $\alpha\beta$-convergent sequence is $\alpha\beta$-bounded. The projection of e_k in σ_1 on u in ${\sigma_1}^*$ is u_k, and, since ${\sigma_1}^* = \sigma_\infty$, e_k is coordinate convergent and projective bounded in σ_1, but is not projective convergent.

(12) A point x such that its projection on every β vector is the limit of the projections of $x^{(n)}$ in α upon that vector is called the *projective limit of $x^{(n)}$ relative to* β, or the $\alpha\beta$-limit of $x^{(n)}$. When $\beta = \alpha^*$, x is the *projective limit* of $x^{(n)}$ in α.

Thus projective limit is based upon the requirement that the limit process and projection should be interchangeable (L.), This limit does not always exist. The fundamental result that the projective limit exists whenever β is normal follows from Theorem I, which is proved by K.T., 198, for a normal α and is proved in our theorem for every α. The projective limit coincides with the coordinate limit and may or may not be in α. It is in this respect that our definition of limit differs from that of K.T., 200; for they restrict the limit to the space of the convergent sequences.

(13) It is shown by K.T. what types of sequence are (projective) convergent in σ_r $(r > 1)$ and in σ_∞; we obtain the corresponding results in ϕ, σ_1 and σ. It is necessary for this to introduce the idea of strong projective convergence. For a fixed x in α and u in β, $f(u) = \Sigma u_k x_k$ defines a function in β. We may regard the sequence of projections of $x^{(n)}$ in α on β vectors as a sequence of functions $f^{(n)}(u)$ defined in β. If these projective functions are uniformly convergent whenever u is restricted to a projective bounded set U in β, *i.e.* if $|f^{(p)}(u) - f^{(q)}(u)| \leqslant \epsilon$ for $p, q \geqslant N(\epsilon, U)$, we say that $x^{(n)}$ is *strongly projective convergent relative to* β. When $\beta = \alpha^*$, $x^{(n)}$ is strongly projective convergent in α and our definition coincides in this case with that of K.T. for strong convergence (p. 201).

(14) A point x is called the *strong projective limit of* $x^{(n)}$ *relative to* β if, whenever u is in a projective bounded set U in β, the projections of $x^{(n)}$ on u converge uniformly to the projection of x on u. When $\beta = \alpha^*$, x is the *strong projective limit of* $x^{(n)}$ *in* α. This definition differs from that of K.T., 201, in that x is not restricted to α. It will be shown that the strong projective limit always exists.

It is proved by K.T., 219, that in σ_1 (projective) convergence implies strong (projective) convergence. We prove this result for ϕ and σ and also show what type of sequence is strongly projective convergent in σ_r $(r > 1)$. The corresponding theorem for σ_∞ is given by K.T., 219. It will be shown that the set of projective limits and the set of strong projective limits of sequences in α relative to β are linear spaces, and these will be found in certain cases.

The concept of closure is used by K.T., 197, to classify spaces in which their limit always exists. With our extended notion of limit, it is used to classify spaces which contain every limit, the definition being based upon position and not upon existence.

If $x^{(n)}$ is $\alpha\beta$-convergent, the matrix (x_{nk}) transforms every sequence in β into a convergent sequence, and so the problem of determining which sequences are $\alpha\beta$-convergent is that of determining which matrices, with rows in α, transform β into Γ. We are thus considering the transformation of a whole class of sequences and not that of a particular sequence. For example, the rows of a K matrix are in σ_1 and it transforms Γ into itself. Similarly, if $x^{(n)}$ is $\alpha\beta$-bounded, (x_{nk}) transforms β into σ_∞. It will be shown that in certain cases, and always when β is complete, it transforms every projective bounded set U in β into a collectively bounded set in σ_∞. If $x^{(n)}$ is strongly $\alpha\beta$-convergent, (x_{nk}) transforms every projective bounded set in β into a collectively bounded and collectively convergent set in Γ.

1. *Projective convergence.*

We use α and β to denote sequence spaces such that $\phi \leqslant \beta \leqslant \alpha^*$.

From the Cauchy test for sequences of numbers we see that a necessary and sufficient condition for the $\alpha\beta$-convergence of $x^{(n)}$ in α is that to every u in β and to every $\epsilon > 0$ there corresponds a positive number $N(\epsilon, u)$, such that

$$\left| \sum_k u_k(x_k^{(p)} - x_k^{(q)}) \right| \leqslant \epsilon$$

for every $p, q \geqslant N$.

When β is normal a better test is given by

THEOREM 1. (Proved by K.T., 198 and 201, for a normal α.) *When β is normal, a necessary and sufficient condition for $x^{(n)}$ in α to be $\alpha\beta$-convergent is that to every u in β and to every $\epsilon > 0$ corresponds an $N(\epsilon, u)$ such that $\sum_k |u_k(x_k^{(p)} - x_k^{(q)})| \leqslant \epsilon$ for every $p, q \geqslant N$.*

It follows from the above that the condition is sufficient; we have to prove it is necessary. Suppose that $x^{(n)}$ is $\alpha\beta$-convergent and that the condition is not satisfied. Then there is an $\epsilon > 0$, a u in β and a sequence of pairs of indices p_i, q_i such that

$$\sum_{k=1}^{\infty} |u_k(x_k^{(p_i)} - x_k^{(q_i)})| > \epsilon. \tag{1}$$

Determine N_1 so that

$$\sum_{k=N_1+1}^{\infty} |u_k(x_k^{(p_1)} - x_k^{(q_1)})| \leqslant \frac{\epsilon}{5}, \tag{2}$$

and therefore

$$\sum_{k=1}^{N_1} |u_k(x_k^{(p_1)} - x_k^{(q_1)})| > \frac{4\epsilon}{5}. \tag{3}$$

Put $v_k = u_k \epsilon_k$, where $|\epsilon_k| = 1$ for every k, and determine the arguments of ϵ_k for $1 \leqslant k \leqslant N_1$ so that $v_k(x_k^{(p_1)} - x_k^{(q_1)})$ is real and not negative. Then

$$\sum_{k=1}^{N_1} v_k(x_k^{(p_1)} - x_k^{(q_1)}) = \sum_{k=1}^{N_1} |u_k(x_k^{(p_1)} - x_k^{(q_1)})| > \frac{4\epsilon}{5};$$

and so, whatever the arguments of ϵ_k for $k > N_1$ may be, we have

$$\begin{aligned} \left| \sum_{k=1}^{\infty} v_k(x_k^{(p_1)} - x_k^{(q_1)}) \right| &= \left| \sum_{k=1}^{N_1} v_k(x_k^{(p_1)} - x_k^{(q_1)}) + \sum_{k=N_1+1}^{\infty} v_k(x_k^{(p_1)} - x_k^{(q_1)}) \right| \\ &\geqslant \left| \sum_{k=1}^{N_1} |u_k(x_k^{(p_1)} - x_k^{(q_1)})| - \sum_{k=N_1+1}^{\infty} |v_k(x_k^{(p_1)} - x_k^{(q_1)})| \right|, \end{aligned}$$

and $|v_k| = |u_k|$ for every k, so that (2) and (3) give

$$\left| \sum_{k=1}^{\infty} v_k (x_k^{(p_1)} - x_k^{(q_1)}) \right| > \frac{3\epsilon}{5} > \frac{\epsilon}{5}.$$

Now $a\beta$-convergence implies coordinate convergence, and from coordinate convergence we can determine i_2 so that

$$\sum_{k=1}^{N_1} |v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| \leqslant \frac{\epsilon}{5}. \tag{4}$$

Now choose $N_2 > N_1$ so that

$$\sum_{k=N_2+1}^{\infty} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| \leqslant \frac{\epsilon}{5}. \tag{5}$$

From (1) it follows that

$$\sum_{k=1}^{N_2} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| > \frac{4\epsilon}{5}. \tag{6}$$

Determine the arguments of ϵ_i for $N_1 + 1 \leqslant i \leqslant N_2$ so that

$$\begin{aligned} \sum_{k-N_1+1}^{N_2} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) &= \sum_{k=N_1+1}^{N_2} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| \\ &= \sum_{k=1}^{N_2} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| - \sum_{k=1}^{N_1} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| \\ &> \frac{4\epsilon}{5} - \frac{\epsilon}{5} = \frac{3\epsilon}{5}, \end{aligned}$$

from (4) and (6). Hence

$$\begin{aligned} &\left| \sum_{k=1}^{\infty} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) \right| \\ &\geqslant \left| \sum_{k=N_1+1}^{N_2} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) \right| - \left| \sum_{k=1}^{N_1} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) \right| - \left| \sum_{k=N_2+1}^{\infty} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) \right| \\ &\geqslant \left| \sum_{k=N_1+1}^{N_2} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) \right| - \sum_{k=1}^{N_1} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| - \sum_{k=N_2+1}^{\infty} |u_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})})| \\ &> \frac{3\epsilon}{5} - \frac{\epsilon}{5} - \frac{\epsilon}{5} = \frac{\epsilon}{5}, \end{aligned}$$

from (4) and (5).

β is normal and $|v_k| = |u_k|$, so that v is in β and

$$\left| \sum_{k=1}^{\infty} v_k (x_k^{(p_{i_2})} - x_k^{(q_{i_2})}) \right| > \frac{\epsilon}{5},$$

whatever the arguments of v_k may be for $k > N_2$.

We thus construct a point v in β so that

$$\left| \sum_{k=1}^{\infty} v_k (x_k^{(p_i)} - x_k^{(q_i)}) \right| > \frac{\epsilon}{5}$$

for an infinite sequence of pairs of indices p_i, q_i. From the generalized Cauchy test it follows that $x^{(n)}$ is not $\alpha\beta$-convergent in such a case. The condition is therefore necessary when β is normal.

In this proof the only property of β that is assumed is that, if it contains u, then it contains v whenever $|v_k| = |u_k|$ for every k. We show that such a space must be normal, *i.e.* that it must contain y whenever $|y_k| \leqslant |u_k|$. For let $u_k = a_k + ib_k$, where a_k, b_k are real, and let $\bar{u}_k = a_k - ib_k$. Then $|u_k| = |\bar{u}_k|$ and the point $\bar{u}$, constructed with coordinates $\bar{u}_k$, is in β. It follows from the definitions of sequence spaces that the points $u+\bar{u}$ and $u-\bar{u}$ are in β and so (a_k) and (b_k) are in β. Now let $|y_k| \leqslant |u_k|$ and put $y_k = c_k + id_k$. Then $c_k^2 + d_k^2 \leqslant a_k^2 + b_k^2$. Determine f_k so that $c_k^2 + f_k^2 = a_k^2 + b_k^2$ and construct the point $z = (c_k + if_k)$. Then $|z_k| = |u_k|$ and so z is in β and therefore (c_k) is in β. Next determine g_k so that $g_k^2 + d_k^2 = a_k^2 + b_k^2$ and construct the point $x = (g_k + id_k)$. Then $|x_k| = |u_k|$ and x is in β and therefore (d_k) is in β. Therefore $(c_k + id_k)$ is in β, *i.e.* y is in β and β is normal.

Projective limit.

Definition. A point x is called the *projective limit* of $x^{(n)}$ in α relative to β or $\alpha\beta$-lim $x^{(n)}$ when

(i) $\Sigma x_k u_k$ is absolutely convergent for every u in β;

(ii) $\lim_{n\to\infty} \sum_{k=1}^{\infty} x_k^{(n)} u_k = \sum_{k=1}^{\infty} x_k u_k$ for every u in β.

For $u = e_k$ we have $\lim_{n\to\infty} x_k^{(n)} = x_k$, and so, if $\alpha\beta$-lim $x^{(n)} = x$, then c-lim $x^{(n)} = x$. It follows from (1) that c-limits of $\alpha\beta$-cgt sequences are considered as possible $\alpha\beta$-limits only if they are in β^*. However, a sequence in α may be $\alpha\beta$-cgt and the c-limit may not satisfy conditions (i) and (ii). For example, let

$$x_k^{(n)} = \left(\frac{n}{n+1}\right)^k.$$

Then $x^{(n)}$ is in σ_1. Let u be any point in D and let $u_{k+1} = -u_k$ for $k > p$. Then

$$\sum_k u_k x_k^{(n)} = \sum_{k=1}^{p} u_k \left(\frac{n}{n+1}\right)^k + u_{p+1} \sum_{r=1}^{\infty} \left(\frac{n}{n+1}\right)^{p+r} (-1)^{r+1}$$

$$= \sum_{k=1}^{p} u_k \left(\frac{n}{n+1}\right)^k + u_{p+1} \left(\frac{n}{n+1}\right)^{p+1} \frac{n+1}{2n+1}.$$

Therefore $$\lim_{n\to\infty} \sum_k u_k x_k^{(n)} = \sum_{k=1}^{p} u_k + \tfrac{1}{2} u_{p+1}.$$

Hence $x^{(n)}$ is $\sigma_1 D$-cgt and condition (1) is not satisfied by the c-limit $x_k = 1$.

Consider next the sequence $x^{(n)} = e_n$. It is $\sigma_1\,\Gamma$-cgt and in this case we have $x_k = 0$ for every k and $\sum_k x_k^{(n)} = 1$ for every n, so that (ii) is not satisfied.

When $\beta = \phi$ we have the following result.

THEOREM 2. *$a\phi$-cgce coincides with c-cgce and $a\phi$ limit is coextensive with the c-limits of c-cgt sequences in a.* (L.)

Let $x^{(n)}$ in a be c-cgt, let u be in ϕ, and let $u_k = 0$ for $k > p$. Then if c-lim $x^{(n)} = x$,

$$\lim_{n\to\infty} \sum_{k=1}^{p} x_k^{(n)} u_k = \sum_{k=1}^{p} x_k u_k.$$

Hence every c-cgt sequence in a is $a\phi$-cgt and the c-limit is the $a\phi$-limit of the sequence.

When we adapt K.T., 199, to this extended conception of limit, we have

THEOREM 3. *When β is normal, the c-limit of every $a\beta$-cgt sequence is the $a\beta$-limit of that sequence.*

From Theorem 1 it follows that, given any u in β and any $\epsilon > 0$,

$$\sum_{k=1}^{\infty} |u_k(x_k^{(p)} - x_k^{(q)})| \leqslant \epsilon$$

for $p, q \geqslant N(\epsilon, u)$. Therefore, for every m,

$$\sum_{k=1}^{m} |u_k(x_k^{(p)} - x_k^{(q)})| \leqslant \epsilon$$

for $p, q \geqslant N$. Fix q and increase p. From c-cgce it follows that

(i) $$\sum_{k=1}^{m} |u_k(x_k - x_k^{(q)})| \leqslant \epsilon$$

for $q \geqslant N$ and every m; and so

$$\text{(ii)} \qquad \sum_{k=1}^{\infty} |u_k(x_k - x_k^{(q)})| \leqslant \epsilon$$

for $q \geqslant N$. From (i)

$$\sum_{k=1}^{m} |u_k x_k| \leqslant \epsilon + \sum_{k=1}^{m} |u_k x_k^{(q)}|.$$

Because $x^{(q)}$ is in α and u is in $\beta \leqslant \alpha^*$,

$$\lim_{m\to\infty} \sum_{k=1}^{m} |u_k x_k^{(q)}| = \sum_{k=1}^{\infty} |u_k x_k^{(q)}|.$$

Hence $\sum_{k=1}^{\infty} |u_k x_k|$ is cgt and x is in β^*.

From (2) we have

$$\left| \sum_k u_k(x_k - x_k^{(q)}) \right| \leqslant \sum_k |u_k(x_k - x_k^{(q)}| \leqslant \epsilon$$

for $q \geqslant N(\epsilon, u)$. Hence

$$\lim_{q\to\infty} \sum_k u_k x_k^{(q)} = \sum_k u_k x_k \quad \text{and} \quad \alpha\beta \lim x^{(n)} = x.$$

The next theorem is an extension of K.T., § 3, Satz 2, 198.

THEOREM 4. *If $\alpha \geqslant \phi$, every sequence in β^* is the $\alpha\beta$-limit of its sections.* (L.)

Let $x = (x_k)$ be in β^* and let $x^{(n)} = (x_1, x_2, \ldots, x_n, 0, 0 \ldots)$. Then $x^{(n)}$ is in $\phi \leqslant \alpha$. For u in β, $\Sigma |u_k x_k|$ is cgt and $\lim_{n\to\infty} \sum_{k=1}^{n} u_k x_k = \sum_{k=1}^{\infty} u_k x_k$. Therefore $\alpha\beta\text{-lim } x^{(n)} = x$.

Example. If $\alpha \geqslant \phi$, every point in α^{**} is the projective limit of its sections in α.

2. *Sequence spaces and projective limits.*

The c-limit of an $\alpha\beta$-cgt sequence may or may not be in α. It has, for example, been shown that $\left\{\left(\frac{n}{n+1}\right)^k\right\}$ is σ_1 D-cgt and that the c-limit is not in σ_1. Similarly, $\alpha\beta$-limits are not always in α; thus, the ϕ sequences $x^{(n)}$, where $x_k^{(n)} = 1$ for $k \leqslant n$ and $x_k^{(n)} = 0$ for $k > n$, are $\phi\sigma_1$-cgt.

The $\phi\sigma_1$-limit, $x_k = 1$, is not in ϕ. In order to investigate the problems of existence and position of these limits we make the following definitions.

Definition. If, with a given definition of cgce and limit, the c-limit of every cgt sequence in α is the limit of that sequence, then α is *regular* under the given cgce. (L.)

Definition. A space α is *cgce closed* under a given definition of cgce when the c-limit of every cgt sequence in α is itself in α. The given cgce does not then lead out of α. (L.)

Definition. A space α is *limit closed* under a given definition of cgce when the limit of every cgt sequence in α is itself in α. (L.)

The derived sets of α under $\alpha\beta$-cgce.

Definition.

(i) The set of c-limits of $\alpha\beta$-cgt sequences in α is denoted by $D_\beta(\alpha)$;

(ii) The set of $\alpha\beta$-limits of $\alpha\beta$-cgt sequences in α is denoted by $d_\beta(\alpha)$. (L.)

When $\beta = \alpha^*$, the derived sets are denoted by $D(\alpha)$ and $d(\alpha)$ respectively. Evidently $D_\beta(\alpha) \geqslant d_\beta(\alpha)$. If α is cgce-closed under $\alpha\beta$-cgce, then α is limit closed under $\alpha\beta$-cgce, but the converse does not necessarily hold. For example, if $\sigma_1 D$-limit $x^{(n)} = x$, then x is in $D^* = \sigma_1$ from the definition of projective limit and so σ_1 is limit closed under $\sigma_1 D$-cgce. It has been shown that $\left\{\left(\frac{n}{n+1}\right)^k\right\}$ is $\sigma_1 D$-cgt and that the c-limit $x_k = 1$ is not in σ_1.

It is easily shown that when α contains ϕ, $D_\beta(\alpha)$ and $d_\beta(\alpha)$ are sequence spaces. Each contains the origin, which is the c-limit and $\alpha\beta$-limit of the sequence $(1/n, 0, 0, \ldots)$; and it is readily verified that the other two conditions are satisfied.

THEOREM 5. *When β is normal, α is regular under $\alpha\beta$-cgce.*

This follows immediately from Theorem 3, and we have

$$D_\beta(\alpha) = d_\beta(\alpha) \leqslant \beta^*.$$

COROLLARY 1. *When $\alpha \geqslant \phi$ and β is normal, Theorem 4 gives*

$$D_\beta(\alpha) = d_\beta(\alpha) = \beta^*.$$

Example. If $\alpha \geqslant \phi$, $D_\phi(\alpha) = d_\phi(\alpha) = \sigma$, and so σ is the only space containing ϕ which is closed relative to ϕ.

COROLLARY 2. *a^* is complete and therefore normal, and so every a is regular under projective cgce and $D(a) = d(a) \leqslant a^{**}$.*

From Corollary 1, if $a \geqslant \phi$, $D(a) = d(a) = a^{**}$.

COROLLARY 3. *From Corollary 1 we obtain the following results:*

$$D_{\sigma_\infty}(\phi) = d_{\sigma_\infty}(\phi) = \sigma_\infty{}^* = \sigma_1,$$

$$D_{\sigma_2}(\phi) = d_{\sigma_2}(\phi) = \sigma_2{}^* = \sigma_2,$$

$$D_{\sigma_1}(\phi) = d_{\sigma_1}(\phi) = \sigma_1{}^* = \sigma_\infty,$$

$$D_{\sigma_2}(\sigma_1) = d_{\sigma_2}(\sigma_1) = \sigma_2{}^* = \sigma_2,$$

$$D_{\sigma_1}(\sigma_2) = d_{\sigma_1}(\sigma_2) = \sigma_1{}^* = \sigma_\infty,$$

$$D_{\sigma_1}(\sigma_1) = d_{\sigma_1}(\sigma_1) = \sigma_1{}^* = \sigma_\infty.$$

From Corollary 2 we obtain

$$D(\Gamma) = d(\Gamma) = \Gamma^{**} = \sigma_1{}^* = \sigma_\infty,$$

$$D(C) = d(C) = C^{**} = \sigma_\infty,$$

$$D(D) = d(D) = D^{**} = \sigma_\infty,$$

$$D(Z) = d(Z) = Z^{**} = \sigma_\infty.$$

The results of Theorems 6, 7, 8, which follow, are slightly different from K.T., 198, 201.

THEOREM 6. *Every complete a is cgce closed and limit closed under projective cgce.*

Every complete space contains ϕ. Theorem 5, Corollary 2, gives

$$D(a) = d(a) = a^{**} = a.$$

THEOREM 7. *When $a \geqslant \phi$ and is normal, a^* is closed in both senses under a^* a-cgce.*

For a satisfies the condition $\phi \leqslant a \leqslant a^{**}$ and we may consider a^* a-cgce. Since a^* is complete and contains ϕ, Theorem 5, Corollary 1, gives $D_a(a^*) = d_a(a^*) = a^*$.

Example. $a = Z$ gives σ_1 closed in both senses under σ_1 Z-cgce.

THEOREM 8. *If $a \geqslant \phi$ and is limit closed under $a\beta$-cgce, then a is complete and, in fact, $a = \beta^*$,*

Let x be in α^{**} and construct the sections $x^{(n)} = (x_1, \ldots, x_n, 0, \ldots)$. For u in $\beta \leqslant \alpha^*$ $\lim_{n\to\infty} \sum_{k=1}^{n} u_k x_k = \sum_{k=1}^{\infty} u_k x_k$. Therefore $\alpha\beta\text{-}\lim x^{(n)} = x$ and x is in α. Therefore $\alpha^{**} \leqslant \alpha$. For all sequence spaces $\alpha^{**} \geqslant \alpha$, and therefore $\alpha = \alpha^{**}$ and α is complete.

Now let y be in β^* and construct the sections $y^{(n)}$. For u in β,

$$\lim_{n\to\infty} \sum_{k=1}^{n} u_k y_k = \sum_{k=1}^{\infty} u_k y_k.$$

Therefore $\alpha\beta\text{-}\lim y^{(n)} = y$. Therefore y is in α and so $\beta^* \leqslant \alpha$.

From $\phi \leqslant \beta \leqslant \alpha^*$ we have $\beta^* \geqslant \alpha$ and so $\alpha = \beta^*$.

Example. If $\alpha \geqslant \phi$ and is limit closed under projective cgce, then α is complete.

3. *Bounded sets.*

It has been proved (K.T., 202) that if X is a bounded set in α and U a bounded set in α^*, then xu is bounded for x in X and u in U. Their method gives a more general result.

THEOREM 9. *If* (i) $\phi \leqslant \beta \leqslant \alpha^*$, (ii) $\mu \geqslant \alpha$, (iii) $\phi \leqslant \mu \leqslant \beta^*$, (iv) β *is cgce-closed under* $\beta\mu$*-cgce,* (v) *the set* X *in* α *is* $\alpha\beta$*-bounded,* (vi) *the set* U *in* β *is* $\beta\mu$*-bounded, then* ux *is bounded for* x *in* X *and* u *in* U.

LEMMA. *If* $x^{(n)}$ *in* α *is* $\alpha\beta$*-bounded,* $\Sigma|c_k|$ *cgt, then* $y^{(n)} = \sum_{i=1}^{n} c_i x^{(i)}$ *is* $\alpha\beta$*-cgt.*

For u in β, $|x^{(n)}u| \leqslant R(u)$ for every u. Given $\epsilon > 0$, determine p so that $\sum_{i=p+1}^{\infty} |c_i| \leqslant \frac{\epsilon}{R}$. Then, if in $m \geqslant n > p$,

$$|u(y^{(m)} - y^{(n)}| = \left| u \sum_{i=n+1}^{m} c_i x^{(i)} \right| = \left| \sum_{i=n+1}^{m} c_i u x^{(i)} \right| \leqslant \sum_{i=n+1}^{m} |c_i|\,|ux^{(i)}| \leqslant \epsilon.$$

By the Cauchy Test, $y^{(n)}$ is $\alpha\beta$-cgt. Because β contains e_k, $|x_k^{(n)}| \leqslant m_k$ for every n and therefore $\sum_{n=1}^{\infty} |c_n x_k^{(n)}|$ is cgt. If c-limit $y^{(n)} = y$, then $y_k = \sum_{n=1}^{\infty} c_n x_k^{(n)}$, and we then write $y = \sum_{n=1}^{\infty} c_n x^{(n)}$.

Proof of Theorem 9. Suppose it is not true. Then we can determine $x^{(1)}$ in X and $u^{(1)}$ in U so that

$$|u^{(1)}x^{(1)}| > 1 + \epsilon. \tag{1}$$

Since X is $\alpha\beta$-bounded,

$$|u^{(1)}x| \leqslant s_1 \quad \text{for every } x \text{ in } X. \tag{2}$$

Since U is $\beta\mu$-bounded,

$$|ux^{(1)}| \leqslant r_1 \quad \text{for every } u \text{ in } U. \tag{3}$$

If the integers p_k are such that $p_k > p_{k-1}$ and $p_0 = 1$, then $\sum_{k=1}^{\infty} \frac{1}{2^{p_{k-1}}}$ is cgt; and, if $u^{(k)}$ is a sequence in U, $y^{(n)} = \sum_{k=1}^{n} \frac{1}{2^{p_{k-1}}} u^{(k)}$ is $\beta\mu$-cgt, by the lemma.

Determine the positive integer p_1 so that

$$\frac{1}{2^{p_1-1}} \leqslant \frac{\epsilon}{r_1}. \tag{4}$$

Since β is cgce closed under $\beta\mu$-cgce, if c-limit $y^{(n)} = y$, y is in β and

$$y = \sum_{k=1}^{\infty} \frac{1}{2^{p_{k-1}}} u^{(k)} = u^{(1)} + \frac{1}{2^{p_1}} u^{(2)} + \frac{1}{2^{p_2}} u^{(3)} + \ldots.$$

Hence

$$yx^{(1)} = u^{(1)}x^{(1)} + \frac{1}{2^{p_1}} u^{(2)}x^{(1)} + \frac{1}{2^{p_2}} u^{(3)}x^{(1)} + \ldots,$$

so that

$$|yx^{(1)}| > |u^{(1)}x^{(1)}| - r_1\left[\frac{1}{2^{p_1}} + \frac{1}{2^{p_2}} + \ldots\right] \quad \text{by (3)}$$

$$> (1+\epsilon) - \epsilon \quad \text{from (1) and (4).}$$

Consequently $|yx^{(1)}| > 1$, however the integers p_k are determined for $k \geqslant 2$.

By our hypothesis, we can determine $x^{(2)}$ in X and $u^{(2)}$ in U so that

$$|x^{(2)}u^{(2)}| > 2^{p_1} \quad (s_1 + 2 + \epsilon). \tag{5}$$

Then we have

$$|x^{(2)}u| \leqslant r_2 \quad \text{for every } u \text{ in } U, \tag{6}$$

and

$$|xu^{(2)}| \leqslant s_2 \quad \text{for every } x \text{ in } X. \tag{7}$$

Determine $p_2 > p_1$ so that

$$\frac{1}{2^{p_2-1}} \leqslant \frac{\epsilon}{r_2}. \tag{8}$$

Then

$$yx^{(2)} = u^{(1)}x^{(2)} + \frac{1}{2^{p_1}} u^{(2)}x^{(2)} + \frac{1}{2^{p_2}} u^{(3)}x^{(2)} + \ldots.$$

Therefore

$$|yx^{(2)}| \geqslant \frac{1}{2^{p_1}}|u^{(2)}x^{(2)}| - |u^{(1)}x^{(2)}| - \sum_{k=3}^{\infty} \frac{1}{2^{p_{k-1}}}|u^{(k)}x^{(2)}|$$

$$\geqslant (s_1+2+\epsilon) - s_1 - r_2 \sum_{k=3}^{\infty} \frac{1}{2^{p_{k-1}}} \quad \text{by (5), (2) and (6)}$$

$$\geqslant (s_1+2+\epsilon) - s_1 - r_2 \frac{\epsilon}{r_2} \quad \text{by (8)}$$

$$\geqslant 2.$$

We thus construct a point y in β and a sequence $x^{(n)}$ in X such that $|yx^{(n)}| > n$, a contradiction to the fact that $x^{(n)}$ is $\alpha\beta$-bounded.

Example. The conditions are satisfied if μ contains ϕ and is normal. For then $\beta = \mu^*$ (Theorem 8) and μ^* is cgce closed under μ^* μ-cgce (Theorem 7). We may, for example, take $\alpha = \phi$, $\beta = \sigma_1$ and $\mu = Z$.

COROLLARY 1. *If α contains ϕ, we may consider the case in which $\mu = \alpha$, for then $\phi \leqslant \alpha \leqslant \beta^*$. If β is cgce closed under $\beta\alpha$-cgce, then $\beta = \alpha^*$ (Theorem 8). We thus have the result that if* (i) $\alpha \geqslant \phi$, (ii) α^* *is cgce closed under* $\alpha^*\alpha$*-cgce,* (iii) X *is projective bounded in* α, (iv) U *is* $\alpha^*\alpha$*-bounded, then xu is bounded.*

By Theorem 6, these conditions are always satisfied when α is normal and contains ϕ.

COROLLARY 2 *Take $\mu = \alpha^{**}$, for $\phi \leqslant \alpha^{**} \leqslant \beta^*$. Then, by Theorem 8, $\beta = \alpha^{***} = \alpha^*$. Conversely, α^* is complete and cgce closed under α^*-cgce (Theorem 5). We thus obtain the result given by K.T., namely, that if X and U are projective bounded in α and α^* respectively, xu is bounded.*

In general, when X is $\alpha\beta$ bounded and U is $\beta\alpha^{**}$ bounded, xu is not bounded. For example, let $\alpha = \phi$, $\beta = C$, and then $\alpha^{**} = \phi$. The sequence $x^{(n)}$ in which the n-th coordinate is $1/n$ and the rest zero is ϕC-bounded. The sequence $u^{(m)}$ where $u_k^{(m)} = k^2$ for $1 \leqslant k \leqslant m$ and $u_k^{(m)} = 0$ for $k > m$ is $C\phi$-bounded. In this case $x^{(n)}u^{(n)} = n$.

This corollary shows that if $x^{(n)}$ is projective bounded in α and U is projective bounded in α^*, the matrix (x_{nk}) transforms U into a collectively bounded set in σ_∞.

Example. If X is projective bounded in α and u a point in α^*, then $\sum_k |x_k u_k| \leqslant r(u)$ for every x in X

For let V be the set such that $|v_k| = |u_k|$. The arguments of v_k are arbitrary. Because α^* is complete and therefore normal, V is in α^*. For y in α^{**}

$$\left| \sum_k v_k y_k \right| \leqslant \sum_k |v_k y_k| = \sum_k |u_k y_k|$$

and so V is projective bounded in α^*, and so

$$\left| \sum_k v_k x_k \right| \leqslant r(V)$$

for every x in X and v in V.

Take any x in X and select that particular v in V for which $x_k v_k$ is real and not negative. Then $\sum_k |v_k x_k| \leqslant r$, *i.e.* $\sum_k |u_k x_k| \leqslant r$.

COROLLARY 3. *Take $\mu = \beta^*$. When β is closed under β-cgce, it is complete. (Theorem 7, example.) Conversely, every complete β satisfies this condition. (Theorem 5.) Hence, if β is complete, X $\alpha\beta$-bounded, U projective bounded in β, then xu is bounded.*

There is no similar result when there are no restrictions on β. To show this take $\alpha = \sigma_1$, $\beta = D$ and then $\beta^* = \sigma_1$. It has been shown that the sequence $\left\{\left(\frac{n}{n+1}\right)^k\right\}$ is $\sigma_1 D$-cgt and therefore $\sigma_1 D$-bounded (see Projective Limit). Let $u^{(n)}$ be the sequence in D where $u_k^{(n)} = 1$ for $1 \leqslant k \leqslant n$ and $u_k^{(n)} = 0$ for $k > n$. Then $u^{(n)}$ is $\sigma_1 D$-bounded and

$$u^{(n} x^{(n)} = \sum_{k=1}^{n} \left(\frac{n}{n+1}\right)^k = n\left[1 - \left(\frac{n}{n+1}\right)^n\right],$$

which increases indefinitely with n.

Example. If β is complete and X is $\alpha\beta$-bounded, then to every u in β corresponds an $r(u)$ so that $\sum_k |x_k u_k| \leqslant r$ for every x in X.

The proof is exactly similar to that in Corollary 2 (example).

Properties of projective bounded sets.

Necessary and sufficient conditions that the set X should be projective bounded in σ_∞ are that it should be collectively bounded, and, in σ_r $(r \geqslant 1)$, that $\sum_k |x_k|^r$ is bounded. These are proved in K.T., 218–219.

THEOREM 10. *The set X is projective bounded in σ if, and only if, $|x_k| \leqslant a_k$.*

The condition is necessary because ϕ contains e_k.

Conversely, if u is in ϕ and $u_k = 0$ for $k > p$,

$$\left| \sum_k u_k x_k \right| \leqslant \sum_{k=1}^{p} |u_k a_k|$$

and so the condition is sufficient.

THEOREM 11. *The set X is projective bounded in ϕ if, and only if,* (i) $|x_k| \leqslant a_k$ *and* (ii) $x_k = 0$ *for* $k > l$ *for every x in X.* (L.)

The conditions are sufficient, because for u in σ we then have

$$\left| \sum_k u_k x_k \right| \leqslant \sum_{k=1}^{p} |u_k a_k|$$

for every x in X.

Condition (i) is necessary because σ contains e_k.

If the condition (ii) is not satisfied, there is a sequence $x^{(n)}$ in X whose lengths increase with n, the length being defined as the index of the last non-zero coordinate. Denote the length of $x^{(n)}$ by l_n and let $u_k = 0$ for $1 \leqslant k \leqslant l_1 - 1$ and $u_{l_1} = 1/x_{l_1}^{(1)}$. Then whatever be u_k for $k > l_1$,

$$\Sigma u_k x_k^{(1)} = u_{l_1} x_{l_1}^{(1)} = 1$$

and

$$\Sigma u_k x_k^{(2)} = u_{l_1} x_{l_1}^{(2)} + \sum_{k=l_1+1}^{l_2} u_k x_k^{(2)}.$$

Put $u_k = 0$ for $l_1 < k \leqslant l_2 - 1$ and choose u_{l_2} so that

$$u_{l_2} x_{l_2}^{(2)} = 2 + \frac{a_{l_1}}{|x_{l_1}^{(1)}|}.$$

Then

$$\left| \sum_k u_k x_k^{(2)} \right| = |u_{l_1} x_{l_1}^{(2)} + u_{l_2} x_{l_2}^{(2)}| \geqslant \left(2 + \frac{a_{l_1}}{|x_{l_1}^{(1)}|}\right) - \frac{a_{l_1}}{|x_{l_1}^{(1)}|}$$

by condition (i), *i.e.* $\left| \sum_k u_k x_k^{(2)} \right| \geqslant 2$, whatever u_k may be, for $k > l_2$.

Proceeding in this way, we construct u in σ so that $\left| \sum_k u_k x_k^{(n)} \right| \geqslant n$, which is contrary to the fact that X is projective bounded in ϕ. The condition (ii) is therefore necessary.

4. *Strong projective cgce.*

Definition. If $x^{(n)}$ in α satisfies the condition that to every $\epsilon > 0$ and to every projective bounded set U in β there corresponds an $N(\epsilon, U)$, so that, for every u in U, $\left| \sum_{k=1}^{\infty} u_k (x_k^{(p)} - x_k^{(q)}) \right| \leqslant \epsilon$ for $p, q \geqslant N$, then $x^{(n)}$ is strongly projective cgt relative to β, or strongly $\alpha\beta$-cgt.

When $\beta = \alpha^*$, $x^{(n)}$ is strongly projective cgt in α or strongly α-cgt, and then the definition is the same as that of K.T. for Strong Cgce (p. 201).

By taking U to consist of one point, it is evident that strong $\alpha\beta$-cgce implies $\alpha\beta$-cgce. The converse is not, in general, true. For example, e_k is ϕC-cgt and e_k is projective bounded in C^* $(C^* = \sigma_1)$ and

$$|e_k(e_p - e_q)| = 1$$

if $p = k$ and $q > k$. The sequence e_k is not, therefore, strongly ϕC-cgt.

If $x^{(n)}$ is strongly $\alpha\beta$-cgt, the matrix (x_{nk}) transforms every projective bounded set in β into a collectively cgt set in Γ.

Strong α-cgce implies strong $\alpha\beta$-cgce, because from $\phi \leqslant \beta \leqslant \alpha^*$ we have $\beta^* \geqslant \alpha^{**}$, and so if U is projective bounded in β, then U is in α^* and is projective bounded in α^*.

The generalized Cauchy Test for Strong Projective Cgce is given by the next theorem.

THEOREM 12. *Necessary and sufficient conditions that the sequence $x^{(n)}$ in α shall be strongly $\alpha\beta$-cgt are that, to every $\epsilon > 0$ and to every projective bounded set U in β, there corresponds $N(\epsilon, U)$, so that, for every u in U,*

$$\sum_k |u_k(x_k^{(p)} - x_k^{(q)})| \leqslant \epsilon$$

for $p, q \geqslant N$.

The condition is sufficient, since it implies the definition. To prove it is necessary; construct the sections $u^{(n)} = (u_1, \ldots, u_n, 0, \ldots)$ of every u in U and consider the set V of points $v^{(n)}$ such that $|v_k^{(n)}| = |u_k^{(n)}|$. The arguments of $v_k^{(n)}$ are arbitrary. The points $v^{(n)}$ are in $\phi \leqslant \beta$.

For y in β^*

$$\left|\sum_k v_k^{(n)} y_k\right| \leqslant \sum_{k=1}^{n} |u_k y_k| \leqslant \sum_{k=1}^{\infty} |u_k y_k|$$

for every n. From Theorem 9, Corollary 2, example,

$$\sum_{k=1}^{\infty} |u_k y_k| \leqslant r(y)$$

for every u in U, and so

$$\left|\sum_k v_k^{(n)} y_k\right| \leqslant r(y)$$

and V is projective bounded in β. Consequently, if $x^{(n)}$ is strongly $\alpha\beta$-cgt,

$$\left|\sum_k v_k^{(n)}(x_k^{(p)} - x_k^{(q)})\right| \leqslant \epsilon$$

for $p, q \geqslant N(\epsilon, V)$. Given any two integers $p, q \geqslant N$, choose the arguments of $v_k^{(n)}$ so that $v_k^{(n)}(x_k^{(p)}-x_k^{(q)})$ is real and not negative. Then we have

$$\sum_{k=1}^{n} |u_k(x_k^{(p)}-x_k^{(q)})| \leqslant \epsilon$$

for every n and every u in U, and so

$$\sum_{k=1}^{\infty} |u_k(x_k^{(p)}-x_k^{(q)})| \leqslant \epsilon$$

for every u in U.

In the spaces σ_r $(r>1)$ and σ_∞, e_k is projective cgt and not strongly projective cgt. This is shown immediately on taking e_k as the projective bounded set in the dual space. In σ_1 projective cgce implies strong projective cgce (and conversely). The proof is in K.T., 219.

THEOREM 13. *Projective cgce in ϕ implies strong projective cgce in ϕ and conversely.*

Let $x^{(n}$ be ϕ-cgt, then $x^{(n)}$ is of bounded length l (Theorem 11). Let U be a projective bounded set in $\phi^* = \sigma$. Then $|u_k| \leqslant a_k$ (Theorem 10). Let $\max a_k$ for $1 \leqslant k \leqslant l$ be m. Given $\epsilon > 0$, we can, from c-cgce, determine N so that

$$\sum_{k=1}^{l} |x_k^{(p)}-x_k^{(q)}| \leqslant \frac{\epsilon}{m}$$

for $p, q \geqslant N$. Then

$$\left| \sum_k u_k(x_k^{(p)}-x_k^{(q)}) \right| = \left| \sum_{k=1}^{l} u_k(x_k^{(p)}-x_k^{(q)}) \right| = m \sum_{k=1}^{l} |x_k^{(p)}-x_k^{(q)}| \leqslant \epsilon$$

for $p, q \geqslant N$. Therefore $x^{(n)}$ is strongly ϕ-cgt.

THEOREM 14. *Projective Cgce in σ implies Strong Projective Cgce in σ and conversely.*

Let U be projective bounded in $\sigma^* = \phi$. Then U is of bounded length l (Theorem 11) and $|u_k| \leqslant a_k$ for u in U. Then if $x^{(n)}$ is σ-cgt,

$$\left| \sum_k u_k(x_k^{(p)}-x_k^{(q)}) \right| \leqslant \sum_{k=1}^{l} a_k |x_k^{(p)}-x_k^{(q)}| \leqslant \epsilon$$

for $p, q \geqslant N(\epsilon, U)$ from c-cgce. Therefore $x^{(n)}$ is strongly σ-cgt.

COROLLARY. *From Theorem 2 and the above, a sequence is strongly $a\phi$-cgt if, and only if, it is c-cgt.*

Strong projective limit.

Definition. The point x is the strong projective limit of $x^{(n)}$ in α relative to β when

(i) $\sum_k |u_k x_k|$ is cgt for every u in β;

(ii) To every projective bounded set U in β and every $\epsilon > 0$ corresponds an $N(\epsilon, U)$ so that, for every u in U

$$\left| \sum_k u_k (x_k^{(n)} - x_k) \right| \leqslant \epsilon$$

for $n \geqslant N$.

This limit may be in α, but is necessarily in β^* by (i). For example, if $x^{(n)} = (a, a^2, \ldots, a^n, 0, \ldots)$ where $0 < a < 1$, and U is projective bounded in σ_∞, and if $x = (a^n)$,

$$\left| \sum_k u_k (x_k^{(n)} - x_k) \right| = \left| \sum_{k=n+1}^{\infty} u_k a^k \right| \leqslant M \sum_{k=n+1}^{\infty} a^k$$

for every u in U, and therefore strong $\phi\sigma_\infty$-limit $x^{(n)} = x$, and x is not in ϕ.

It is evident on taking U to consist of one point, that if strong $\alpha\beta$-limit $x^{(n)} = x$, then $\alpha\beta$-limit $x^{(n)} = x$ and also c-limit $x^{(n)} = x$. When $\beta = \phi$, it follows from Theorems 2 and 14 that if x is the c-limit of $x^{(n)}$ in α, then also strong $\alpha\phi$-limit $x^{(n)} = x$.

We now prove the existence theorem for strong projective limits.

THEOREM 15. *Every α is regular under strong $\alpha\beta$-cgce.*

Let $x^{(n)}$ be strongly $\alpha\beta$-cgt, c-limit $x^{(n)} = x$, and let U be a projective bounded set in β. Then, by Theorem 12,

$$\sum_{k=1}^{\infty} |u_k (x_k^{(p)} - x_k^{(q)})| \leqslant \epsilon$$

for $p, q \geqslant N(\epsilon, U)$. Therefore, for every m,

$$\sum_{k=1}^{m} |u_k (x_k^{(p)} - x_k^{(q)})| \leqslant \epsilon$$

for $p, q \geqslant N$. Fix u and q and increase p. From c-cgce we have

$$\sum_{k=1}^{m} |u_k (x_k - x_k^{(q)})| \leqslant \epsilon \tag{1}$$

for every m. Therefore

$$\sum_{k=1}^{\infty} |u_k(x_k - x_k^{(q)})| \leqslant \epsilon. \tag{2}$$

From (1),

$$\sum_{k=1}^{m} |u_k x_k| \leqslant \epsilon + \sum_{k=1}^{m} |u_k x_k^{(q)}|$$

for every m and every u in U;

$$\sum_{k=1}^{\infty} |u_k x_k^{(q)}|$$

is cgt because u is in β and $x^{(q)}$ in α. Therefore

$$\sum_{k=1}^{\infty} |u_k x_k|$$

is cgt. The set U is arbitrary, and so x is in β^*.

From (2) it follows that strong $\alpha\beta$-limit $x^{(n)} = x$.

COROLLARY 1. *If $x^{(n)}$ is strongly $\alpha\beta$-cgt and if c-limit $x^{(n)} = x$, from Theorem* 13 *strong $\alpha\beta$-limit $x^{(n)} = x$, and so, if U is projective bounded in β, from the definition of limit we have*

$$\left| \sum_k u_k x_k^{(n)} \right| \leqslant \left| \sum_k u_k x_k \right| + \epsilon$$

for $n \geqslant N(\epsilon, U)$.

Because x is in β^*,

$$\left| \sum_k u_k x_k \right| \leqslant r(x)$$

for every u in U. Therefore, for $n \geqslant N$ and every u,

$$\left| \sum_k u_k x_k^{(n)} \right| \leqslant r + \epsilon.$$

For $n < N$,

$$\left| \sum_k u_k x_k^{(n)} \right| \leqslant s_n,$$

since $x^{(n)}$ is in $\alpha \leqslant \alpha^{**} \leqslant \beta^*$. Therefore

$$\left| \sum_k u_k x_k^{(n)} \right| \leqslant R$$

for every n and every u in U.

The matrix (x_{nk}) with rows in α, transforms the set U into a collectively bounded and collectively cgt set in Γ.

COROLLARY 2. *If $x^{(n)}$ is strongly $a\beta$-cgt and $a\beta$-limit $x^{(n)} = x$, then strong $a\beta$-limit $x^{(n)} = x$.*

Sequence spaces and strong projective limits.

Consider the sequence $x_k = 1$ in σ_∞ and construct the sections $x^{(n)}$ of this point. The set e_k is projective bounded in $\sigma_\infty{}^* = \sigma_1$ and

$$|e_k(x^{(p)} - x^{(q)})| = |x_k^{(p)} - x_k^{(q)}| = 1$$

if $p = k$ and $q = k-1$. The sections of $x_k = 1$ are not strongly projective cgt.

THEOREM 16. *When a is ϕ. σ_r $(r \geqslant 1)$ or σ, every point in a is the strong projective limit of its sections.*

Consider each space separately. U denotes a projective bounded set in the dual space and $x^{(n)} = (x_1, x_2, \ldots, x_n, 0, \ldots)$.

(i) In ϕ, let $x_k = 0$ for $k > l$, then $\sum_k u_k(x_k^{(n)} - x_k) = 0$ for $k > l$.

(ii) In σ_1, $\left| \sum_{k=1}^{\infty} u_k(x_k^{(n)} - x_k) \right| = \left| \sum_{k=n+1}^{\infty} u_k x_k \right| \leqslant M \sum_{k=n+1}^{\infty} |x_k|$ for every u,

$\leqslant \epsilon$ for $n > N$.

(iii) In σ_r $(r > 1)$,

$$\left| \sum_{k=1}^{\infty} u_k(x_k^{(n)} - x_k) \right| = \left| \sum_{k=n+1}^{\infty} u_k x_k \right| \leqslant \left(\sum_{k=n+1}^{\infty} |u_k|^s \right)^{1/s} \left(\sum_{k=n+1}^{\infty} |x_k|^r \right)^{1/r} \leqslant \epsilon$$

for $n \geqslant N$, by Hölder's Inequality and because $\left(\sum_{k=1}^{\infty} |u_k|^s \right)^{1/s}$ is bounded for u in U.

(iv) In σ, then U in ϕ is of bounded length and $u_k = 0$ for $k > l$ (Theorem 11),

$$\left| \sum_{k=1}^{l} u_k(x_k^{(n)} - x_k) \right| = 0$$

if $n > l$.

COROLLARY 1. *On considering the sections of x in σ_1 as a sequence in ϕ, it follows that every point in σ_1 is the strong $\phi\sigma_\infty$-limit of its sections.*

COROLLARY 2. *If $\phi < a < \sigma_r$, every point in σ_r is the strong $a\sigma_s$ limit of its sections.*

COROLLARY 3. *If $a \geqslant \phi$, every point in σ is the strong $a\phi$ limit of its sections.*

Closed spaces and derived sets.

It follows from Theorem 15 that, if α is limit closed under strong $\alpha\beta$-cgce, then α is cgce closed under strong $\alpha\beta$-cgce, and conversely. As it is not now necessary to distinguish between the two cases, we shall say that such a space α is closed under strong $\alpha\beta$-cgce. It is evident that if α is limit closed under $\alpha\beta$-cgce, then α is closed under strong $\alpha\beta$-cgce. Hence, from Theorem 6, every complete α is closed under strong projective cgce, and from Theorem 7, when α contains ϕ and is normal, α^* is closed under strong $\alpha^*\alpha$-cgce. For example, σ_1 is closed under strong $\sigma_1 Z$-cgce.

Definition. The set of strong $\alpha\beta$-limits of sequences in α is denoted by $d_\beta'(\alpha)$ and, when $\beta = \alpha^*$, by $d'(\alpha)$.

Then $d_\beta'(\alpha)$ is a linear space if $\alpha \geqslant \phi$. For if U is projective bounded in β, $|u_1| \leqslant r$ for every u in U, since β^* contains e_1 in ϕ; and if $x^{(n)} = (1/n, 0, 0, \ldots)$,

$$\left| \sum_k u_k (x_k^{(p)} - x_k^{(q)} \right| = \left| u_1 \left(\frac{1}{p} - \frac{1}{q} \right) \right| \leqslant r \left(\frac{1}{p} + \frac{1}{q} \right)$$

for every u in U. Therefore $x^{(n)}$ is strongly $\alpha\beta$-cgt, and $d_\beta'(\alpha)$ contains the origin.

Also if strong $\alpha\beta$-limit $x^{(n)} = x$ and strong $\alpha\beta$-limit $y^{(n)} = y$, then strong $\alpha\beta$-limit $(x^{(n)} + y^{(n)}) = x + y$ for

$$|u(x^{(n)} + y^{(n)}) - \overline{x+y}| = |u(x^{(n)} - x) + u(y^{(n)} - y)|,$$

and also strong $\alpha\beta$-limit $rx^{(n)} = rx$, and the result follows.

From the definition we have (i) $d_\beta'(\alpha) \leqslant d_\beta(\alpha) \leqslant D_\beta(\alpha)$; (ii) $d_\beta'(\alpha) \leqslant \beta^*$; (iii) $d'(\alpha) \leqslant \alpha^{**}$. From Theorem 16 we now obtain $d'(\phi) = \phi$, $d'(\sigma_r) = \sigma_r$ $(r \geqslant 1)$, and $d'(\sigma) = \sigma$

From Theorem 16, Corollary 1, and (ii) above,

$$d'_{\sigma_\infty}(\phi) = \sigma_1.$$

From Theorem 16, Corollary 2, and (ii) above,

$$d'_{\sigma_2}(\phi) = \sigma_2 \quad \text{and} \quad d'_{\sigma_2}(\sigma_1) = \sigma_2.$$

From Theorem 16, Corollary 3, and (ii) above, $d_\phi'(\alpha) = \sigma$ when α contains ϕ. Hence σ is the only space containing ϕ which is closed under strong projective cgce relative to ϕ. From (ii) above we also have $d'_{\sigma_1}(\phi) \leqslant \sigma_\infty$, $d'_{\sigma_1}(\sigma_1) \leqslant \sigma_\infty$, $d'_{\sigma_1}(\sigma_2) \leqslant \sigma_\infty$.

5. *Properties of projective cgt sequences.*

It is proved by K.T., 204, that if a^* possesses the property that every point in it is the strong (projective) limit of its sections, then a sequence in a is projective cgt when it is c-cgt and projective bounded. A sequence is projective bounded in σ_r $(r > 1)$ when $\sum_k |x_k|^r$ is bounded, and in σ_∞ when $|x_k^{(n)}|$ is bounded (K.T., 218–219), and so the necessary and sufficient conditions for projective cgce in these spaces are immediately obtained. In σ, c-cgce is sufficient (Theorem 2).

Stated in terms of matrices, these results give the following:

(1) If $1/r+1/s = 1$, the matrix (x_{nk}), where $\sum_k |x_{nk}|^r$ is cgt, will transform every sequence u_k, where $\sum_k |u_k|^s$ is cgt, into a cgt sequence if, and only if, (1) $\lim_{n\to\infty} x_{nk} = x_k$ (exists for every k); (2) $\sum_k |x_{nk}|^r \leqslant M$ independently of n.

(2) The matrix (x_{nk}), where $|x_{nk}| \leqslant r_n$ for every k, will transform every sequence u_k, where $\Sigma|u_k|$ is cgt, into a cgt sequence if, and only if, (i) $\lim_{n\to\infty} x_{nk} = x_k$; (ii) $|x_{nk}| \leqslant M$ for every n and k.

(3) The matrix (x_{nk}) will transform every finite sequence into a cgt sequence if, and only if, $\lim_{n\to\infty} x_{nk} = x_k$ for every k.

We have shown that the point $x_k = 1$ in σ_1^* is not the projective limit of its sections. We shall now obtain the conditions for projective cgce in σ_1 and ϕ.

THEOREM 17. *A sequence in ϕ is projective cgt if, and only if, it is c-cgt and of bounded length. A row-finite matrix transforms every sequence into a cgt sequence if, and only if, it is row-bounded and has column limits.*

This follows immediately from Theorem 16 and the K.T. Test.

THEOREM 18. *Necessary and sufficient conditions that a sequence should be projective cgt in σ_1 are that it should be c-cgt and collectively absolutely cgt. The matrix (x_{nk}), where $\sum_k |x_{nk}|$ is cgt, transforms every bounded sequence into a cgt sequence if, and only if,* (i) $\lim_{n\to\infty} x_{nk} = x_k$ *for every* k; (ii) *to every $\epsilon > 0$ corresponds an $m(\epsilon)$ so that $\sum_{k=m+1}^{\infty} |x_{nk}| \leqslant \epsilon$ for every n.*

The conditions are sufficient. Suppose they are satisfied. Given u in σ_∞ where $|u_k| \leqslant M$ and $\epsilon > 0$, determine m so that

$$\sum_{k=m+1}^{\infty} |x_k^{(n)}| \leqslant \frac{\epsilon}{3M}$$

for every n. From c-cgce, determine N so that, for $p, q \geqslant N$,

$$\sum_{k=1}^{m} |u_k(x_k^{(p)} - x_k^{(q)})| \leqslant \tfrac{1}{3}\epsilon.$$

Then

$$\left| \sum_{k=1}^{\infty} u_k(x_k^{(p)} - x_k^{(q)}) \right| \leqslant \sum_{k=1}^{m} |u_k(x_k^{(p)} - x_k^{(q)})| + \sum_{k=m+1}^{\infty} |u_k x_k^{(p)}| + \sum_{k=m+1}^{\infty} |u_k x_k^{(q)}|$$
$$\leqslant \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon = \epsilon$$

for $p, q \geqslant N$.

The result follows from the Cauchy Test.

The conditions are necessary. Suppose $x^{(n)}$ is σ_1-cgt. From Theorem 6, $D(\sigma_1) = d(\sigma_1) = \sigma_1$, and if projective limit $x^{(n)} = x$, then x is in σ_1. Given $\epsilon > 0$, determine m so that

$$\sum_{k=m+1}^{\infty} |x_k| \leqslant \tfrac{1}{2}\epsilon. \tag{1}$$

From Theorem 3, Equation 2, for every u in σ_∞,

$$\sum_{k=1}^{\infty} |u_k(x_k - x_k^{(q)})| \leqslant \tfrac{1}{2}\epsilon$$

for $q \geqslant N(\epsilon, u)$. Put $u_k = 0$ for $1 \leqslant k \leqslant m$ and $|u_k| = 1$ for $k > m$. Then

$$\sum_{k=m+1}^{\infty} |x_k - x_k^{(q)}| \leqslant \tfrac{1}{2}\epsilon$$

for $q \geqslant N'$. Therefore

$$\sum_{k=m+1}^{\infty} |x_k^{(q)}| \leqslant \sum_{k=m+1}^{\infty} |x_k| + \tfrac{1}{2}\epsilon \leqslant \epsilon$$

by (1) for $q \geqslant N'$.

We can determine n so that $\sum_{k=n+1}^{\infty} |x_k^{(q)}| \leqslant \epsilon$ for $1 \leqslant q \leqslant N'$, and the result follows.

COROLLARY. *The conditions of this theorem are also the conditions for $\phi\sigma_\infty$-cgce.*

6. Properties of strong projective cgt sequences.

It is shown by K.T., 216, that the conditions for strong (projective) cgce in σ_2 are

$$\sqrt{\left(\sum_k |x_k^{(p)} - x_k^{(q)}|\right)^2} \leqslant \epsilon$$

for $p, q \geqslant N(\epsilon)$, and in σ_∞ (K.T., 219), $|x_k^{(p)} - x_k^{(q)}| \leqslant \epsilon$ for $p, q \geqslant N$ and every k.

We have shown that in σ, c-cgce is sufficient (Theorems 2 and 14), and in ϕ the conditions are c-cgce and bounded length (Theorems 13 and 17). In σ_1 projective cgce implies strong projective cgce, and conversely (K.T., 219), and the conditions are stated in Theorem 18, namely, c-cgce and collective absolute cgce. Suppose α is one of the spaces ϕ, σ_1, σ_2, σ_∞ or σ. The properties of a projective bounded set in α^* are known and the matrix (x_{nk}), with rows in α, will transform every such set into a collectively bounded and collectively cgt set in Γ.

It remains to obtain the corresponding result for σ_r $(r > 1)$.

THEOREM 19. *Necessary and sufficient conditions that $x^{(n)}$ should be strongly projective cgt in σ_r $(r > 1)$ are* (i) $\lim_{n\to\infty} x_k^{(n)} = x_k$ *for every k;* (ii) $\sum_{k=m+1}^{\infty} |x_k^{(n)}|^r \leqslant \epsilon$ *for every n, i.e. $x^{(n)}$ is c-cgt and the series $\sum_k |x_k^{(n)}|^r$ collectively absolutely cgt.*

If $\frac{1}{r} + \frac{1}{s} = 1$, the matrix (x_{nk}), where $\sum_k |x_{nk}|^r$ is cgt, will transform every set U, which is such that $\sum_k |u_k|^s$ is bounded, into a collectively bounded and collectively cgt set if (i) $\lim_{n\to\infty} x_{nk} = x_k$ *for every k;* (ii) *to every $\epsilon > 0$ corresponds an $m(\epsilon)$, so that*

$$\sum_{k=m+1}^{\infty} |x_{nk}|^r \leqslant \epsilon$$

for every n.

The conditions are necessary. We have $\sigma_r^* = \sigma_s$ where $1/r + 1/s = 1$ and $\sigma_r^{**} = \sigma_r$. Let $x^{(n)}$ be strongly projective cgt in σ_r, then strong projective limit $x^{(n)} = x$ is in σ_r. Hence to every projective bounded set U in σ_s and to every $\epsilon > 0$ corresponds $N(\epsilon, U)$, so that

$$\sum_{k=1}^{\infty} |u_k(x_k^{(n)} - x_k)| \leqslant \tfrac{1}{2}\epsilon \qquad (1)$$

for $n \geqslant N$ (Theorem 15).

Let the set U' be constructed in the following manner. To every u' in U' corresponds $x^{(p)}$ such that either $|u_k'| = 0$ or $|u_k'| = |x_k^{(p)}|^{r-1}$. Then $|u_k'|^s \leqslant |x_k^{(p)}|^{rs-s} = |x_k^{(p)}|^r$. Therefore U' is in σ_s.

Since $x^{(n)}$ is projective bounded in σ_r,

$$\sum_k |x_k^{(n)}|^r \leqslant M \tag{2}$$

for every n. Therefore $\sum_k |u_k'|^s \leqslant M$ for every u' in U', and U' is projective bounded in σ_s. Since x is in σ_r, we can determine m so that

$$\left(\sum_{k=m+1}^{\infty} |x_k|^r\right)^{1/r} \leqslant \frac{\epsilon}{2M^{1/s}}. \tag{3}$$

In (1) put $u_k = 0$ for $1 \leqslant k \leqslant m$ and $|u_k| = |x_k^{(p)}|^{r-1}$ for $k > m$. Then

$$\sum_{k=m+1}^{\infty} |x_k^{(p)}|^{r-1} |x_k^{(n)} - x_k| \leqslant \tfrac{1}{2}\epsilon$$

for $n \geqslant N'$ and every p. Therefore

$$\sum_{k=m+1}^{\infty} |x_k^{(p)}|^{r-1} |x_k^{(n)}| \leqslant \sum_{k=m+1}^{\infty} |x_k^{(p)}|^{r-1} |x_k| + \tfrac{1}{2}\epsilon. \tag{4}$$

By Hölder's Inequality,

$$\begin{aligned}
\sum_{k=m+1}^{\infty} |x_k^{(p)}|^{r-1} |x_k| &\leqslant \left(\sum_{k=m+1}^{\infty} |x_k|^r\right)^{1/r} \left(\sum_{k=m+1}^{\infty} |x_k^{(p)}|^{sr-s}\right)^{1/s} \\
&\leqslant \frac{\epsilon}{2M^{1/s}} \left(\sum_{k=m+1}^{\infty} |x_k^{(p)}|^r\right)^{1/s} \quad \text{by (3)} \\
&\leqslant \frac{\epsilon}{2M^{1/s}} M^{1/s} \quad \text{by (2) for every } p \\
&\leqslant \tfrac{1}{2}\epsilon.
\end{aligned}$$

From (4), for every p and every $n \geqslant N'$,

$$\sum_{k=m+1}^{\infty} |x_k^{(p)}|^{r-1} |x_k^{(n)}| \leqslant \epsilon,$$

Since N' is independent of p, we may put $p = n \geqslant N'$. Therefore, for $n \geqslant N'$,

$$\sum_{k=m+1}^{\infty} |x_k^{(n)}|^r \leqslant \epsilon,$$

and the result follows.

The conditions are sufficient. When they are satisfied,

$$\sum_k |u_k|^s \leqslant M \tag{1}$$

for every u in U. Given $\epsilon > 0$, determine m so that

$$\left(\sum_{k=m+1}^{\infty} |x_k^{(n)}|^r \right)^{1/r} \leqslant \frac{\epsilon}{3M^{1/s}} \tag{2}$$

for every n. Then using Hölder's Inequality,

$$\left| \sum_{k=1}^{m} u_k (x_k^{(p)} - x_k^{(q)}) \right| \leqslant \left(\sum_{k=1}^{m} |u_k|^s \right)^{1/s} \left(\sum_{k=1}^{m} |x_k^{(p)} - x_k^{(q)}|^r \right)^{1/r}$$

$$\leqslant M^{1/s} \left(\sum_{k=1}^{m} |x_k^{(p)} - x_k^{(q)}|^r \right)^{1/r} \quad \text{for every } u \text{ in } U \text{ from (1)}$$

$$\leqslant \tfrac{1}{3}\epsilon \quad \text{for } p, q \geqslant N \text{ from c-cgce.}$$

Also

$$\left| \sum_{k=m+1}^{\infty} u_k x_k^{(n)} \right| \leqslant \left(\sum_{k=m+1}^{\infty} |u_k|^s \right)^{1/s} \left(\sum_{k=m+1}^{\infty} |x_k^{(n)}|^r \right)^{1/r}$$

$$\leqslant M^{1/s} \frac{\epsilon}{3M^{1/s}} = \tfrac{1}{3}\epsilon$$

using (1) and (2). Then for $p, q \geqslant N$,

$$\left| \sum_{k=1}^{\infty} u_k (x_k^{(p)} - x_k^{(q)}) \right| = \left| \sum_{k=1}^{m} u_k (x_k^{(p)} - x_k^{(q)}) + \sum_{k=m+1}^{\infty} u_k x_k^{(p)} - \sum_{k=m+1}^{\infty} u_k x_k^{(q)} \right|$$

$$\leqslant \sum_{k=1}^{m} |u_k (x_k^{(p)} - x_k^{(q)})| + \left| \sum_{k=m+1}^{\infty} u_k x_k^{(p)} \right| + \left| \sum_{k=m+1}^{\infty} u_k x_k^{(q)} \right|$$

$$\leqslant \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon = \epsilon.$$

Therefore $x^{(n)}$ is strongly projective cgt in σ_r.

Distance convergence.

The sequence spaces are metrical spaces and distance, $d(x, y)$, may be defined in ϕ and σ_∞ by $d(x, y) = \max |x_k - y_k|$ and in σ_r $(r \geqslant 1)$ by

$$d(x, y) = \left(\sum_k |x_k - y_k|^r \right)^{1/r}.$$

It is shown by K.T., 216 and 218, that in σ_1, σ_2 and σ_∞, d-cgce implies strong (projective) cgce, and conversely. We show that this is true in σ_r $(r > 1)$.

THEOREM 20. *In σ_r $(r>1)$, d-cgce implies strong projective cgce, and conversely.*

Let $x^{(n)}$ be d-cgt in σ_r and U a projective bounded set in σ_s such that

$$\left(\sum_k |u_k|^s\right)^{1/s} \leqslant M$$

for every u in U. Determine N so that

$$\left(\sum_k |x_k^{(p)}-x_k^{(q)}|^r\right)^{1/r} \leqslant \frac{\epsilon}{M}$$

for $p, q \geqslant N$. Then

$$\left|\sum_k u_k(x_k^{(p)}-x_k^{(q)})\right| \leqslant \left(\sum_k |u_k|^s\right)^{1/s}\left(\sum_k |x_k^{(p)}-x_k^{(q)}|^r\right)^{1/r} \quad \text{by Hölder's Inequality}$$

$$\leqslant \epsilon \quad \text{for } p, q \geqslant N \text{ and every } u \text{ in } U.$$

Therefore $x^{(n)}$ is strongly projective cgt in σ_r.

Conversely, let $x^{(n)}$ be strongly projective cgt in σ_r. By Theorem 19, we can determine N so that

$$\sum_{k=N+1}^{\infty} |x_k^{(n)}|^r \leqslant \frac{\epsilon^r}{2^{r+1}}. \tag{1}$$

From c-cgce, determine N' so that

$$\sum_{k=1}^{N} |x_k^{(p)}-x_k^{(q)}|^r \leqslant \tfrac{1}{2}\epsilon^r \tag{2}$$

for $p, q \geqslant N'$. Using Minkowski's Inequality, we have, for $p, q \geqslant N'$,

$$\left(\sum_{k=N+1}^{\infty} |x_k^{(p)}-x_k^{(q)}|^r\right)^{1/r} \leqslant \left(\sum_{k=N+1}^{\infty} |x_k^{(p)}|^r\right)^{1/r} + \left(\sum_{k=N+1}^{\infty} |x_k^{(q)}|^r\right)^{1/r}$$

$$\leqslant \frac{\epsilon}{2^{(r+1)/r}} + \frac{\epsilon}{2^{(r+1)/r}} \quad \text{by (1)}$$

$$= \frac{\epsilon}{2^{1/r}},$$

so that

$$\sum_{k=N+1}^{\infty} |x_k^{(p)}-x_k^{(q)}|^r \leqslant \tfrac{1}{2}\epsilon^r. \tag{3}$$

Then

$$\sum_{k=1}^{\infty} |x_k^{(p)}-x_k^{(q)}|^r = \sum_{k=1}^{N} |x_k^{(p)}-x_k^{(q)}|^r + \sum_{k=N+1}^{\infty} |x_k^{(p)}-x_k^{(q)}|^r$$

$$\leqslant \tfrac{1}{2}\epsilon^r + \tfrac{1}{2}\epsilon^r = \epsilon^r \quad \text{by (2) and (3).}$$

Therefore $$\left(\sum_{k=1}^{\infty} |x_k^{(p)} - x_k^{(q)}|^r\right)^{1/r} \leqslant \epsilon$$

for $p, q \geqslant N'$, and hence $x^{(n)}$ is d-cgt.

Bibliography.

S. Banach, *Théorie des opérations linéaires* (Warsaw, 1932).

P. Dienes, *The Taylor series* (Oxford University Press, 1931).

G. Köthe and O. Toeplitz, "Lineare Raume mit unendlichvielen Koordinaten und Ringe unendlicher Matrizen", *Journal für Math.*, 171 (1934), 193–226.

P. Dienes, Lectures on the transformation of sequences, at Birkbeck College, London University.

Birkbeck College,
London.

ON THE GEOMETRY OF NUMBERS IN SOME NON-CONVEX REGIONS

By L. J. MORDELL.

[Received 25 July, 1941.—Read 26 June, 1941.]

1. *Introduction**.

Let $\phi(\xi, \eta)$ be a real function of two real variables ξ, η. It is an interesting and important question to investigate the value, say μ, of the minimum or lower bound of $\phi(\xi, \eta)$ for integer values of ξ, η. Since the set of values of $\phi(\xi, \eta)$ for all ξ, η is unaltered by linear substitutions in ξ, η, with integer coefficients and determinant unity, it is to be expected that μ is invariant for such linear substitutions. When $\phi(\xi, \eta)$ is a homogeneous function of dimension p in ξ, η and p is a positive integer, it is naturally suggested that μ can be expressed algebraically in terms of the usual invariants of $\phi(\xi, \eta)$. The results when $p = 2$ are classic, while those for $p = 3$ were found very recently (1) by me. When p is not an integer, $\phi(\xi, \eta)$ will not in general have any algebraic invariants, but a substitute can, nevertheless, be introduced by considering the question from a slightly different point of view.

Let L be a lattice of determinant $\Delta > 0$ defined by

$$x = \alpha\xi + \beta\eta, \quad y = \gamma\xi + \delta\eta, \tag{1.1}$$

where $\alpha, \beta, \gamma, \delta$ are real constants,

$$\alpha\delta - \beta\gamma = \Delta, \tag{1.2}$$

* I am greatly indebted to Dr. Mahler and Dr. Davenport for having verified my proofs, and for many useful and valuable comments in reading my manuscript. Dr. Mahler has drawn the diagrams given as well as several others of use to me in my work. He also calculated several tables which were very helpful to me in the first stages of this work in connection with § 8. I should like to thank Dr. Rado for reading the proof sheets.

and ξ, η run through all integer values. Then Δ is invariant for linear substitutions in ξ, η of determinant unity.

Let $f(x, y)$ be a homogeneous function of x and y of dimension $p > 0$. Suppose also that $f(x, y)$ is defined for all x and y, and does not vanish except at the origin O, $(x, y) = (0, 0)$. Regions $\Re$ symmetrical with respect to O can be associated with $f(x, y)$ as follows. Let k be a number independent of α, β, γ, δ, and so of the lattice L. One such region is

$$\left|f(|x|, |y|)\right| \leqslant k\Delta^{\frac{1}{2}p}. \tag{1.3}$$

If, further, $f(x, y) = \pm f(-x, -y)$, another is

$$|f(x, y)| \leqslant k\Delta^{\frac{1}{2}p}. \tag{1.4}$$

The question arises whether such regions contain points of L other than O. It is easy to find crude values of k for which this is so. When the region $\Re$ is convex and closed, an admissible value of k is easily deduced from Minkowski's classic theorem stated as

LEMMA 1. *A convex closed region symmetrical with respect to O and of area at least 4Δ, contains a point of L other than O.*

When the region $\Re$ is not convex, crude values of k are easily deduced by making $\Re$ contain simple convex regions, e.g. parallelograms; and in particular the parallelogram of area 4Δ,

$$|x|+|y| \leqslant (2\Delta)^{\frac{1}{2}}. \tag{1.5}$$

which contains a point of L other than O. It is not so easy, however, to find the best possible value of k. Until very recently, no best possible results were known except in the well-known case when $f(x, y) = xy$, more usually taken as $f(x, y) = x^2+xy-y^2$, and due to Korkine and Zolotareff (2) and proved by Markoff (3). Davenport (4) gave a result when $f(x, y)$ is a complicated function arising incidentally in his discussion of the minimum of the product of three ternary linear forms. Recently, I (1) gave the best possible results

$$|xy(x+y)| \leqslant \sqrt{(\tfrac{1}{7})}\,\Delta^{\frac{3}{2}}, \tag{1.6}$$

$$|x^3+y^3| \leqslant \sqrt[4]{(\tfrac{2}{2}\tfrac{7}{3})}\,\Delta^{\frac{3}{2}}. \tag{1.7}$$

In both these cases, $\Re$ is an infinite non-convex region.

Another region obviously presenting itself for investigation is

$$|x|^p+|y|^p \leqslant k\Delta^{\frac{1}{2}p}, \tag{1.8}$$

which is non-convex when $0 < p < 1$.

When $p \geqslant 1$, (1.8) defines a convex region for which the best possible result has been investigated by Minkowski (5). He notes that the question is equivalent to finding the least value of Δ such that a lattice of determinant Δ has none of its points other than O as inner points of the region

$$|x|^p + |y|^p \leqslant 1. \tag{1.9}$$

He proves that these values of Δ arise from lattices which are such that the points given by $(\xi, \eta) = (1, 0), (0, 1), (1, -1)$ lie on the boundary of (1.9). Thus the problem is reduced to that of finding the minimum value of

$$\Delta = \alpha\delta - \beta\gamma, \tag{1.10}$$

where the parameters $\alpha, \beta, \gamma, \delta$ are subject to the conditions

$$|\alpha|^p + |\gamma|^p = 1, \quad |\beta|^p + |\delta|^p = 1, \quad |\alpha - \beta|^p + |\gamma - \delta|^p = 1. \tag{1.11}$$

This is a question in calculus, but not an easy one. He states the result in the form of a geometrical hypothesis, which does not appear to have been proved, and gives the conjectured *minimum* value of Δ.

When $0 < p < 1$, the best possible value of k has not been considered either by Minkowski or other writers, who undoubtedly have been deterred by the non-convexity of the region (1.8). However, by methods related to those used in the proofs of (1.6), (1.7), I have discovered the existence for some non-convex regions of better crude results and also best possible results. The latter are comparable in simplicity and generality with those of Minkowski on convex regions, and lead to the same type of minimum problem. It will be noted that the final results are given explicitly and that their truth depends on a minimum value of Δ occurring for specified lattices, just as in Minkowski's work on convex regions when a minimum value of Δ was given. It is of interest that parallelograms, whose vertices and middle points all lie on the boundary of the region considered, play a fundamental part in the investigation, as in the proofs of (1.6), (1.7).

2. *The function $f(x, y)$.*

Let a function $f(x, y)$ satisfy the following conditions:

I. *$f(x, y)$ is defined for $x \geqslant 0$, $y \geqslant 0$, and is symmetrical in x, y and homogeneous of dimension $p > 0$, that is*

$$f(x, y) = f(y, x), \quad f(x, y) = x^p f(1, y/x). \tag{2.1}$$

II. *For* $x \geqslant 0$, $y \geqslant 0$, $f(x, y)$ *is continuous and* $f(x, y) \geqslant 0$. *Also* $f(1, y)$ *and, for* $0 \leqslant y \leqslant 1$, $f(1+y, 1-y)$ *are respectively steadily increasing and decreasing functions of* y. *Hence also for* $x \geqslant 0$, $y \geqslant 0$, $f(x, y)$ *is an increasing function of either variable when the other is constant.*

III. *The region defined by* $x \geqslant 0$, $y \geqslant 0$, $f(x, y) \geqslant f(1, 1)$ *is convex*†.*

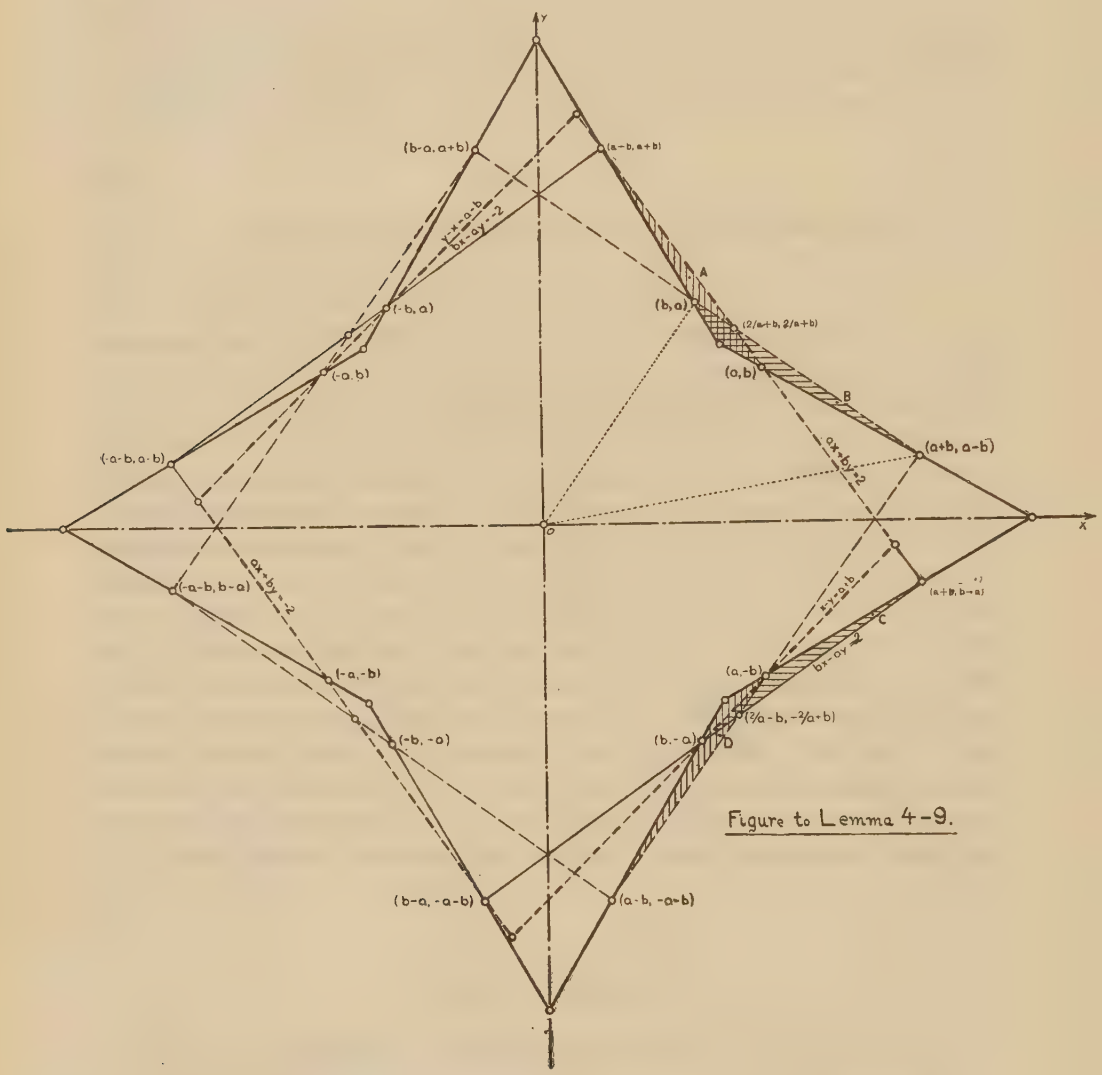


Figure to Lemma 4-9.

* Davenport points out that the monotony conditions in II are a consequence of III when the region terminates in the axes or has them as asymptotes.

† The two figures will be useful in studying this paper. It suffices to take a linear boundary as in Theorem 4.

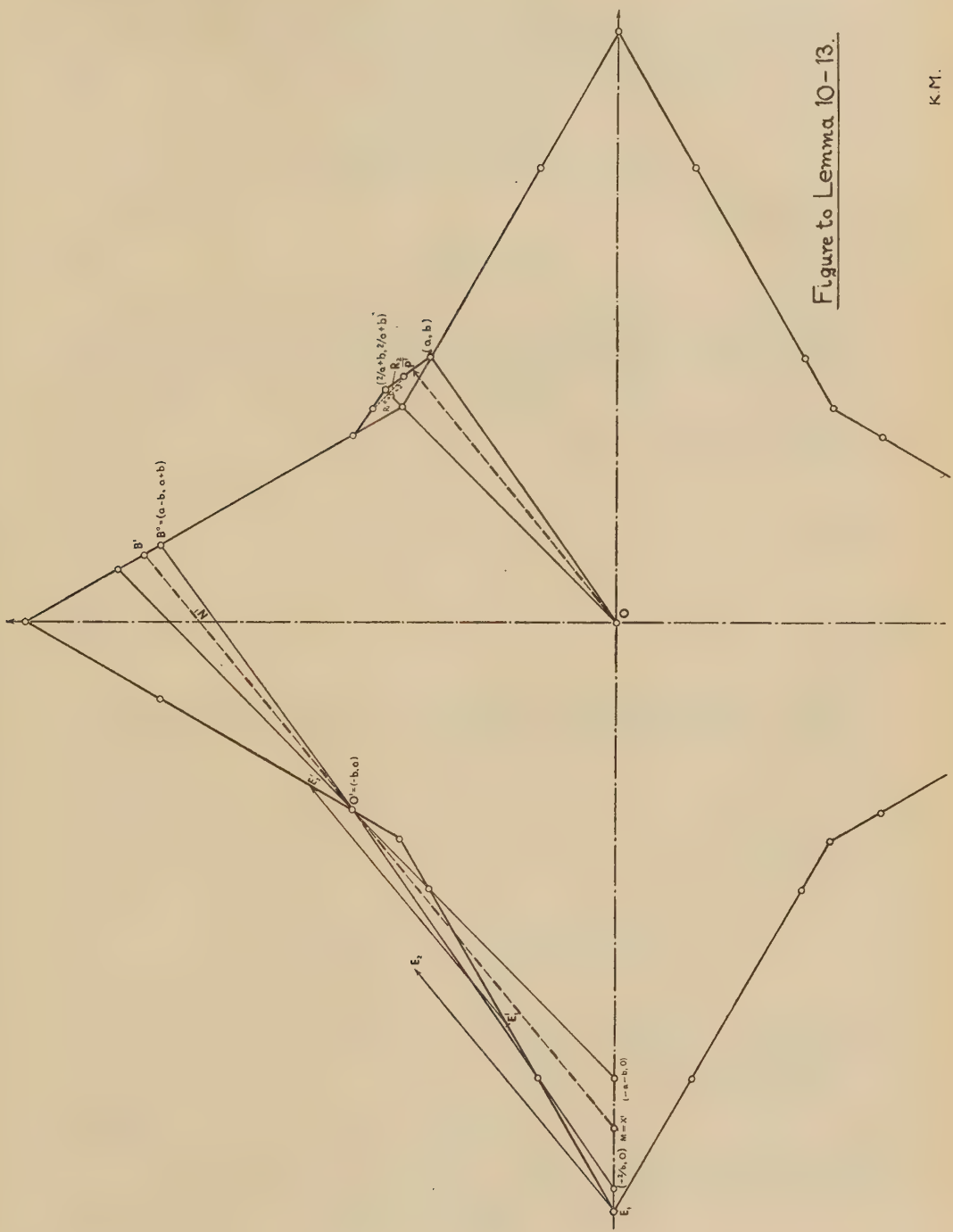


Figure to Lemma 10–13.
K.M.
O
B'
B''=(a-b, a+b)
N
O''=(-b, a)
(a, b)
(½a+b, ½a+b)
(-½b, 0) M=N'
(-a-b, a)

There is no loss of generality in supposing p to be unity since we need only replace $f(x, y)$ by $|f(x, y)|^{1/p}$; we do so except in the applications.

The convexity and symmetry conditions applied to the points (x, y), (y, x) and their midpoint, show that if $x \geqslant 0$, $y \geqslant 0$, then

$$f(x, y) \leqslant f\left(\frac{x+y}{2}, \frac{x+y}{2}\right) = \left(\frac{x+y}{2}\right) f(1, 1). \tag{2.2}$$

In particular from $x = 1$, $y = 0$,

$$f(1, 1) \geqslant 2f(1, 0). \tag{2.3}$$

The equality sign occurs only when $f(x, y) \equiv x+y$, a case that may be omitted hereafter. The inequality (2.3) is significant only when $f(1, 0) \neq 0$, and then there is no loss of generality in taking $f(1, 0) = 1$, and this will be done, but only in the proofs.

Some fundamental numbers associated with $f(x, y)$ are now defined in

LEMMA 2. *Unique numbers a, b, c with $a > b > 0$ are defined by the equations*

$$f(a+b, a-b) = f(a, b) = cf(1, 1), \tag{2.4}$$

$$a^2+b^2 = 2. \tag{2.5}$$

Further $0 < c < 1$.

For write $b = aj$. Then it suffices to show that a unique number j with $0 < j < 1$ is defined by the equation

$$g(j) = f(1+j, 1-j)-f(1, j) = 0. \tag{2.6}$$

From II, $g(z)$ is obviously a monotone decreasing function of z in $0 < z < 1$. From (2.3),

$$g(0) = f(1, 1)-f(1, 0) \geqslant f(1, 1)(1-2^{-1}) > 0,$$

$$g(1) = f(2, 0)-f(1, 1) < f(1, 1)-f(1, 1) = 0.$$

Since $g(z)$ is a continuous function of z and $g(0)$, $g(1)$ have opposite signs, the result for j follows.

It is clear that

$$j > \tfrac{1}{2}, \tag{2.7}$$

since $g(\frac{1}{2}) = f(\frac{3}{2}, \frac{1}{2})-f(1, \frac{1}{2}) > 0$. It is also clear from (2.5) that

$$a+b < 2. \tag{2.8}$$

From (2.4), (2.2),

$$cf(1, 1) = f(a, b) \leqslant \left(\frac{a+b}{2}\right) f(1, 1).$$

Hence

$$c \leqslant \tfrac{1}{2}(a+b) < 1. \tag{2.9}$$

We now prove a result required later and given in

LEMMA 3.

$$f(1, j) > 2f(1, 0). \tag{2.10}$$

As stated above, we may suppose that $f(1, 0) = 1$. If $f(1, j) \leqslant 2$, then, by (2.6), $f(1+j, 1-j) \leqslant 2$. This is false since $f(1+z, 1-z)$ decreases steadily from $f(1, 1) > 2$ to $f(2, 0) = 2$ as z increases from 0 to 1, and $j < 1$.

The result (2.10) can also be written as

$$2a < cf(1, 1). \tag{2.11}$$

For, by (2.10),

$$f\left(1, \frac{b}{a}\right) > 2, \quad f(a, b) > 2a,$$

and so from (2.4)

$$cf(1, 1) > 2a,$$

whence (2.11).

It is convenient to have another inequality connecting a and c, namely that given by

LEMMA 4.

$$a < 2c. \tag{2.12}$$

For, since $j = b/a > \frac{1}{2}$,

$$cf(1, 1) = f(a, b) > f(a, \tfrac{1}{2}a) = \tfrac{1}{2}af(2, 1) > \tfrac{1}{2}af(1, 1),$$

whence the result.

3. *Results obtained.*

The question of finding values of k independent of α, β, γ, δ such that the region

$$f(|x|, |y|) \leqslant k\Delta^{\frac{1}{2}} \tag{3.1}$$

contains a point of L other than O is now considered.

Since, by (1.5), a point of L other than O satisfies

$$|x|+|y| \leqslant \sqrt{(2\Delta)}, \tag{3.2}$$

it follows from (2.2) that this point satisfies

$$f(|x|, |y|) \leqslant f(1, 1)(\tfrac{1}{2}\Delta)^{\frac{1}{2}}. \tag{3.3}$$

The first stage in the improvement of this result is the derivation of the fundamental

LEMMA 7. *If the region $\mathcal{R}$, defined by*

$$f(|x|, |y|) \leqslant cf(1, 1)(\tfrac{1}{2}\Delta)^{\frac{1}{2}}, \tag{3.4}$$

does not contain in its interior a point of L other than O, then a point of L is contained in at least one of the two following regions: either the finite region external to $\mathcal{R}$, but including its boundary, and bounded by the finite lines joining the points

$$\left((\tfrac{1}{2}\Delta)^{\frac{1}{2}} b, (\tfrac{1}{2}\Delta)^{\frac{1}{2}} a\right), \quad \left((\tfrac{1}{2}\Delta)^{\frac{1}{2}} a, (\tfrac{1}{2}\Delta)^{\frac{1}{2}} b\right)$$

of its boundary to the point

$$\left(\frac{2(\frac{1}{2}\Delta)^{\frac{1}{2}}}{a+b}, \frac{2(\frac{1}{2}\Delta)^{\frac{1}{2}}}{a+b}\right),$$

or the image of this region in the y axis.

This result takes a simpler form if $\Delta = 2$. Since the details throughout are also simpler when $\Delta = 2$, we shall assume, except in the enunciation of the applications, that $\Delta = 2$, an assumption which of course involves no loss of generality.

Next, sufficient conditions can be given under which the crude result (3.3) can be replaced by the sharper inequality (3.4). Thus if it can be shown that the conclusions of Lemma 7 except for the boundary parts of $\mathcal{R}$, lead to a contradiction, then $\mathcal{R}$ contains in its interior or on its boundary between the two points mentioned a point of L other than O. One method of establishing this contradiction leads to the fundamental

THEOREM 1. *Let P be any point on the boundary of $\mathcal{R}$ lying between the points (b, a) and (a, b), the end points included. Let QR be a line equal and parallel to OP and such that the ends Q, R lie on the second quadrant boundary of $\mathcal{R}$ (the existence of such lines is proved in Lemma 10). Then if for all such points P, the parallelogram $OPQR$ has an area not less than 2,*

e.g. if the area increases steadily as P moves from the point (b, a) to the point (c, c), a point of L other than O satisfies

$$f(|x|, |y|) \leqslant cf(1, 1). \tag{3.4}$$

The condition, area $OPQR \geqslant 2$, can be easily expressed in the equivalent form: The line parallel to OP and distant $2/OP$ therefrom has a length not greater than OP intercepted by the second quadrant boundary of $\mathfrak{R}$.

Theorem 1 is easily expressed analytically. For let (x, y) be the coordinates of P and $(-q, r)$ be the coordinates of the middle point of QR. Then Q, R are the points $(-q+\frac{1}{2}x, r+\frac{1}{2}y)$ and $(-q-\frac{1}{2}x, r-\frac{1}{2}y)$. Also the area of the parallelogram $OPQR$ is $qy+rx$. Hence we may suppose that

$$x > 0, \quad y > 0, \quad q > \tfrac{1}{2}x, \quad r > \tfrac{1}{2}y, \tag{3.5}$$

$$f(q-\tfrac{1}{2}x, r+\tfrac{1}{2}y) = cf(1, 1), \tag{3.6}$$

$$f(q+\tfrac{1}{2}x, r-\tfrac{1}{2}y) = cf(1, 1), \tag{3.7}$$

$$f(x, y) = cf(1, 1); \tag{3.8}$$

and we require for all the points P lying on the boundary of $\mathfrak{R}$ between the points (b, a) and (c, c), the end points included, that

$$qy+rx \geqslant 2. \tag{3.9}$$

Conditions can now be given under which (3.4) is the best possible result. The method of proof of Theorem 1 leads to

THEOREM 2. *Suppose now in Theorem 1 that $P(x, y)$ is a point such that the area of the parallelogram $OPQR$ is 2, and that Q, R are the boundary points of $\mathfrak{R}$ mentioned there. Then a lattice L is generated by OP, OQ and contains R; and takes the form*

$$X = x\xi+(-q-\tfrac{1}{2}x)\eta, \quad Y = y\xi+(r-\tfrac{1}{2}y)\eta, \tag{3.10}$$

where ξ, η run through all integer values.

If L has no points except O in the interior of $\mathfrak{R}$, i.e. if

$$f(|X|, |Y|) \geqslant cf(1, 1); \tag{3.11}$$

and if this holds for all such points P, then (3.4) is the best possible result.

We can then call such a lattice L a *critical lattice*. When Theorem 1 holds, the only critical lattices arise as in Theorem 2.

In particular, when P is (a, b) or (b, a), the possible critical lattices become

$$X = a\xi + b\eta, \quad Y = b\xi - a\eta, \tag{3.12}$$

$$X = b\xi + a\eta, \quad Y = a\xi - b\eta, \tag{3.13}$$

Also (3.12) is a critical lattice if

$$f(|a\xi + b\eta|, |b\xi - a\eta|) \geqslant cf(1, 1) \tag{3.14}$$

for all integers ξ, η except 0, 0. Sufficient conditions for this are given in Lemma 14 when $\mathfrak{R}$ is finite.

When, in Theorem 2, there is only one other such point P, beside these two points (b, a), (a, b), it must by symmetry be the point (c, c). It is then clear that $q = r = 1/c$, since q, r for given x, y are uniquely determined. The possible critical lattice now is

$$X = c\xi + \left(-\tfrac{1}{2}c - \frac{1}{c}\right)\eta, \quad Y = c\xi + \left(-\tfrac{1}{2}c + \frac{1}{c}\right)\eta. \tag{3.15}$$

By Lemma 5, this is different from (3.12), (3.13). The condition (3.11) now becomes

$$f\left(\left|\xi + \left(-\tfrac{1}{2} - \frac{1}{c^2}\right)\eta\right|, \left|\xi + \left(-\tfrac{1}{2} + \frac{1}{c^2}\right)\eta\right|\right) \geqslant f(1, 1) \tag{3.16}$$

for all integers ξ, η except 0, 0.

When this condition is satisfied, there will by symmetry be another critical lattice which is the image of (3.15) in the Y axis.

A few words may now be said about the relation between my results and those of Minkowski for convex bodies. In the best possible result (3.4), the determinant $\Delta = 2$ in (1.2) is supposed given and c is the least number independent of $\alpha, \beta, \gamma, \delta$ such that (3.4) is satisfied by a point other than O of every lattice of determinant 2.

The question can be put in Minkowski's form by considering a variable Δ and the fixed region

$$f(|x|, |y|) \leqslant 1. \tag{3.17}$$

We have then to find the least value of Δ such that every lattice of determinant Δ has a point other than O in (3.17). This least value of Δ, assuming that the result of Theorem 1 holds, is given by

$$cf(1, 1)(\tfrac{1}{2}\Delta)^{\frac{1}{2}} = 1, \tag{3.18}$$

where c is still defined by (2.4), (2.5).

In the particular case of Theorem 1 when the area of the parallelogram increases steadily, this means that the least value of

$$\Delta = \alpha\delta - \beta\gamma$$

has been found when the parameters α, β, γ, δ satisfy the conditions

$$f(|\alpha|, |\gamma|) = 1, \quad f(|\beta|, |\delta|) = 1,$$

$$f(|\alpha+\beta|, |\gamma+\delta|) = 1,$$

and also $\alpha > 0$, $\gamma > 0$, $\beta < 0$, $\delta > 0$, $\alpha+\beta < 0$. This is clear on taking P, Q, R to be the points (α, γ), (β, δ), $(\alpha+\beta, \gamma+\delta)$ respectively. The minimum of Δ occurs when the parameters correspond to the critical lattices.

When the function $f(x, y)$ does not involve a parameter, the question whether Theorem 2 applies or not can be definitely answered, i.e. settled in a finite time by numerical approximation. The application of Theorem 2 is fairly straightforward when $\mathfrak{B}$, the first quadrant boundary of $\mathfrak{R}$, consists of straight lines or circular arcs, and quite manageable when $\mathfrak{B}$ consists of a cubic or quartic curve. I give five entirely new results in Theorems 3 to 7. In Theorems 3, 4, an arbitrary parameter is involved.

In Theorem 3, $\mathfrak{B}$ is a circular arc cutting the coordinate axes at any given angle. No unusual features arise and there are exactly two critical lattices.

In Theorem 4, $\mathfrak{B}$ consists of two straight lines forming a reentrant quadrilateral with the x, y axes, so that $\mathfrak{R}$ is a regular star-shaped octagon of fixed acute vertical angle 2θ, say. When $60° < 2\theta < 90°$, there are exactly two critical lattices. When $2\theta = 60°$, there are exactly four critical lattices. When $2\theta < 60°$, Theorem 2 does not apply. Presumably for successive ranges of values of 2θ, new critical lattices exist.

In Theorem 5, $f(x, y) = x^p + y^p$, $0 < p < 1$. The best possible results are proved for $p = \frac{1}{2}, \frac{1}{3}$. For the other values of p, it seems probable that the improved crude result (3.4) holds. The truth of this would follow from the conjecture* that the curve

$$[(x+1)^p - 1][(y+1)^p - 1] = (1-x^p)(1-y^p)$$

cuts every line parallel to $x+y = 0$ in at most two points in the first quadrant. I can prove this for $p = \frac{1}{2}$. But, in fact, the result is required

* This has since been proved by Watson (**6**). It follows at once that (8.4) is the best possible result for $1 > p \geqslant \frac{4}{5}$. See the addendum to § 8. Watson's further results (**7**) enable me to replace the $\frac{4}{5}$ by 0·3295 ....

for only a special line, and I can prove this for $0 < p < \frac{1}{3}$. But now a new difficulty arises since the possible critical lattices (3.12), (3.13) may not really be critical lattices.

No parameters are involved in Theorems 6, 7 wherein $\mathcal{R}$ is an infinite region. Thus in Theorem 6, $f(x, y) = xy(x+y)$. Here the possible critical lattices (3.12), (3.13) turn out to have points on the coordinate axes, but no other points, in the interior of $\mathcal{R}$. We still have a significant best possible theorem in the sense that if the dimensions of $\mathcal{R}$ are diminished in a ratio less than one, these lattices (3.12), (3.13) have no points in the region other than those on the coordinate axes.

In Theorem 7, $f(x, y) = xy(x^2+y^2)$, or what comes to the same thing, x^4-y^4. There are now exactly four critical lattices.

As a postscript, I prove the classical result for $f(x, y) = xy$ when there are a continuous infinity of critical lattices.

The present paper suggests a number of questions, e.g. what are the corresponding results when Theorem 2 does not apply, as in Theorem 4, when $2\theta < 60°$; or, again, what happens in connection with Theorem 5? Further, can Theorem 6 be sharpened for lattices different from (3.12), (3.13)? Again it seems not unlikely that theorems similar to Theorem 2 exist for other simple types of non-convex regions. An obvious case for investigation is when $\mathcal{R}$ is a region of a type akin to (1.7), in which the first quadrant boundary of $\mathcal{R}$ has a single point of inflexion.

These remarks make clear that the geometry of numbers in non-convex domains offers far more interesting possibilities than the theory for convex regions, and has now become a most promising field for further investigation and future research.

4. *First stage in proof of Theorem* 1.

We consider more closely the region $\mathcal{R}$ defined by (3.4), namely,

$$f(|x|, |y|) \leqslant cf(1, 1), \tag{4.1}$$

and assume that no points of the lattice L other than O satisfy (4.1) with the inequality sign.

Since the region is symmetrical with respect to each of the x, y coordinate axes, it suffices to describe the part of $\mathcal{R}$ in the first quadrant. Its boundary is $f(x, y) = cf(1, 1)$, and meets the axes of x, y at the points $\big(cf(1, 1), 0\big)$ and $\big(0, cf(1, 1)\big)$, or at infinity according as $f(1, 0) \neq 0$ or $= 0$. The region contained between this boundary, and also including it,

and the line joining these points is conveniently denoted by

$$\left[\left(cf(1,1),0\right);\ \left(0,cf(1,1)\right)\right]$$

(and a similar notation is used throughout for such regions). It is convex and has no points lying in $\mathfrak{R}$.

There are two obvious squares whose vertices and middle points of sides all lie on the boundary of $\mathfrak{R}$.

The first square has its vertices at

$$(a+b,\ a-b),\quad (-a+b,\ a+b),\quad (-a-b,\ -a+b),\quad (a-b,\ -a-b),$$

and the middle points of its sides at

$$(b,\ a),\quad (-a,\ b),\quad (-b,\ -a),\quad (a,\ -b).$$

With the help of (2.5) the equations of its sides are seen to be

$$bx+ay=2,\quad ax-by=-2,\quad \text{etc.} \tag{4.2}$$

All the points of the square except two subsets are internal points of $\mathfrak{R}$. One subset contains the boundary points of $\mathfrak{R}$ lying on the boundary joining $(a+b,\ a-b)$ and $(b,\ a)$, or on the boundary joining $(-a+b,\ a+b)$ and $(-a,\ b)$, together with the other points of the regions

$$[(a+b,\ a-b);\ (b,\ a)]\quad \text{and}\quad [(-a+b,\ a+b);\ (-a,\ b)],$$

which are of course external to $\mathfrak{R}$. The other subset is the image of the first subset in the origin.

The second square referred to above is the image of the first in the x axis, and is also given by an interchange of a and b above.

Each of these squares defines a lattice containing the vertices and midpoints of the sides of the square. These lattices are given by

$$x=a\xi+b\eta,\quad y=b\xi-a\eta, \tag{4.3}$$

$$x=b\xi+a\eta,\quad y=a\xi-b\eta. \tag{4.4}$$

It is obvious that these lattices have no other points except O in the interior of the corresponding square. This is also an immediate consequence of

LEMMA 5. *The triangle whose vertices are* O, $(b,\ a)$, $(a+b,\ a-b)$ *cannot contain, in addition to* O, *two points of* L *non-collinear with* O, *unless* L *is one of the lattices* (4.3), (4.4).

For suppose that two such points exist and call them P_1, P_2. Then the area of the triangle $OP_1P_2 \geqslant 1$, since twice this area is an integral multiple of $\Delta = 2$. But clearly

$$2 \text{ area } OP_1P_2 \leqslant a(a+b)-b(a-b) = a^2+b^2 = 2,$$

by (2.5). Hence the equality sign holds; therefore P_1, P_2 coincide with (b, a) and $(a+b, a-b)$, and then L becomes the lattice (4.4).

LEMMA 6. *A point of L lies in one of the two regions* $[(a, -b); (b, -a)]$, $[(a-b, a+b); (a, b)]$; *and also in one of the two regions* $[(a, -b); (b, -a)]$, $[(b, a); (a+b, a-b)]$.

The first part follows from Lemma 1 applied to the parallelogram

$$|x-y| \leqslant a+b, \quad |ax+by| \leqslant 2, \tag{4.5}$$

since the first two regions and their images in the origin are the parts of the parallelogram not internal to $\mathfrak{R}$. For, writing $X = x-y$, $Y = ax+by$, so that $\dfrac{\partial(X, Y)}{\partial(x, y)} = a+b$, we see that the area of the parallelogram is

$$8 = \frac{4 \times 2(a+b)}{\partial(X, Y)/\partial(x, y)}.$$

Also the points $(a, -b)$, $(b, -a)$ lie on $x-y = a+b$, and the points $(a-b, a+b)$, (a, b) on $ax+by = 2$. We show, further, that the vertices of the parallelogram (4.5) are inner points of $\mathfrak{R}$. Thus the two sides

$$x-y = a+b, \quad ax+by = 2,$$

meet at

$$x = \frac{a^2+b^2}{a+b} + b < a+b,$$

$$y = \frac{a^2+b^2}{a+b} - a = \frac{b(b-a)}{a+b}.$$

Clearly $0 > y > b-a$, since $a > b > 0$, and so this vertex is an inner point of $\mathfrak{R}$.

For the vertex given by

$$x-y = -a-b, \quad ax+by = 2,$$

we have

$$x = \frac{a^2+b^2}{a+b} - b > 0,$$

and hence it lies in the first quadrant. By the convexity Condition III,

it is an inner point of $\mathfrak{R}$, since two points of the side $ax+by=2$ already lie on the boundary of $\mathfrak{R}$. This completes the proof of the first part.

The result for the second pair of regions follows in the same way by considering the parallelogram

$$|x-y| \leqslant a+b, \quad |bx+ay| \leqslant 2. \tag{4.6}$$

There next follows the very important

LEMMA 7 (*principal lemma*). *If the region $\mathfrak{R}$ defined by*

$$f(|x|, |y|) \leqslant cf(1, 1) \tag{4.7}$$

does not contain in its interior a point of L other than O, then a point of L is contained in at least one of the two following regions: either the finite region external to $\mathfrak{R}$ but including its boundary and bounded by the finite line joining the points (b, a), (a, b) of its boundary to the point $\left(\frac{2}{a+b}, \frac{2}{a+b}\right)$; or the region which is the image of this in the y axis.

The lemma is obviously true if a point of L lies in either the curvilinear triangle whose vertices are at (b, a), (a, b), $\left(\frac{2}{a+b}, \frac{2}{a+b}\right)$, or the image of this triangle in the y axis. We may assume, then, for the proof of the lemma, that this does not occur. Then a point of L does not lie in either of the regions $[(a, b); (b, a)]$ and $[(a, -b); (b, -a)]$. Hence, by Lemma 6, a point of L must lie in each of the regions $[(a-b, a+b); (a, b)]$ and $[(b, a); (a+b, a-b)]$. By a similar argument from the regions obtained from those of Lemma 6 on taking their images in the x axis, a point of L must also lie in each of the regions $[(a+b, -a+b); (b, -a)]$ and $[(a, -b); (a-b, -a-b)]$. Call these four points of L, in this order, A, B, C, D. These points are all distinct, for if B e.g. coincided with A, there would be a point of L in the triangle vertices at (b, a), (a, b), $\left(\frac{2}{a+b}, \frac{2}{a+b}\right)$. Since there are only four sets of residues of the sets (ξ, η) (mod 2) for integers ξ, η, two of the five points O, A, B, C, D must give the same set of residues. Hence one of the midpoints of the lines joining O, A, B, C, D in pairs must be a point of L. This is impossible for the pairs OA, OB, OC, OD, e.g. for B, from

LEMMA 8. *If B, a point of L, is contained in the region*

$$[(b, a); (a+b, a-b)],$$

then the midpoint of OB is an inner point of $\mathfrak{R}$.

Since the equation of the line joining the points (b, a), $(a+b, a-b)$ is $bx+ay=2$, it suffices to show that the segment $\frac{1}{2}b \leqslant x \leqslant \frac{1}{2}(a+b)$ of the line $bx+ay=1$ does not meet the boundary $f(x, y)=cf(1, 1)$ in any points lying in the first quadrant. At their intersection, if any,

$$f\left(x, \frac{1-bx}{a}\right) = cf(1, 1),$$

i.e.

$$f(ax, 1-bx) = acf(1, 1).$$

It is now proved, however, that, for all x in $\frac{1}{2}b \leqslant x \leqslant \frac{1}{2}(a+b)$,

$$f(ax, 1-bx) < acf(1, 1).$$

For this it suffices, from (2.2), to prove that $\frac{1}{2}(a-b)x+\frac{1}{2} < ac$, i.e. that this holds for the largest value of x, $x=\frac{1}{2}(a+b)$. Hence we have to show that $\frac{1}{4}(a^2-b^2)+\frac{1}{4}(a^2+b^2) < ac$, i.e. that $a < 2c$. This holds by (2.12).

This argument disposes of the midpoints of OA, OB, OC, OD. It shows also that no points of L are inner points of the line OB.

By symmetry, we need consider only the midpoints of AB, AC, AD, BC, say F, G, H, I, in this order. We dispose of them by

LEMMA 9. *The points F, G, H, I cannot be points of L.*

For F we note that the triangles AOF, BOF have areas $=\frac{1}{2}(a^2+b^2) \geqslant 1$, since, e.g., triangle AOF contains non-collinear lattice points O, A, F. The sum of these areas does not exceed the area of the triangle whose vertices are at O and $(a-b, a+b)$, $(a+b, a-b)$, and which is therefore of area

$$\tfrac{1}{2}(a+b)^2-\tfrac{1}{2}(a-b)^2 = 2ab.$$

Hence $2ab \geqslant a^2+b^2$, which is false.

For G, we note, by Lemmas 5 and 8, that G cannot lie in the triangle vertices at O, $(a-b, a+b)$, (a, b), which already contains A, unless L is the lattice (4.3). Similarly, G cannot lie in the triangle whose vertices are at O, $(a+b, -a+b)$, $(b, -a)$, which already contains C, unless L is the lattice (4.3). The result follows now since G clearly lies in the quadrilateral with vertices at O, $(a-b, a+b)$, $(a+b, -a+b)$, $(b, -a)$.

For H, say the point (x, y), clearly from (2.7),

$$0 < x \leqslant a, \quad |y| \leqslant \tfrac{1}{2}(a+b)-\tfrac{1}{2}b = \tfrac{1}{2}a < b.$$

Hence $f(|x|, |y|) \leqslant f(a, \frac{1}{2}a) < f(a, b)$, and so H is an inner point of $\mathcal{R}$.

We dispose finally of the assumption that $I=\frac{1}{2}(B+C)$ is a lattice point by considering the point $I'=\frac{1}{2}(B-C)$, say (x, y), which is then

also a point of L. Clearly

$$0 \leqslant |x| \leqslant \tfrac{1}{2}(a+b)-\tfrac{1}{2}b = \tfrac{1}{2}a, \quad |y| \leqslant a.$$

Then, as above,

$$f(|x|, |y|) \leqslant f(\tfrac{1}{2}a, a) < f(b, a) = cf(1, 1),$$

and so I' is an inner point of $\mathcal{R}$.

This completes the proof of Lemma 7.

There is now no loss of generality in supposing that a point of L, which it has been supposed has none of its points except O lying in $\mathcal{R}$, say $P'(x', y')$, lies in the finite region in the first quadrant external to $\mathcal{R}$, but including its boundary, bounded by the lines joining the points (b, a), $\left(\frac{2}{a+b}, \frac{2}{a+b}\right)$ and (a, b), $\left(\frac{2}{a+b}, \frac{2}{a+b}\right)$. We divide this region into two parts, $\mathcal{R}_1$ and $\mathcal{R}_2$, according as its points satisfy $x+y < 2$ or $x+y \geqslant 2$.

5. *Final stage in proof of Theorem* 1.

LEMMA 10. *There exists a line $Q'R'$ which is equal and parallel to OP' and whose ends Q', R' are in the second quadrant and lie on the boundary of $\mathcal{R}$.*

Suppose, first, that $f(1, 0) \neq 0$. Draw from the point E_1 $(-cf(1, 1), 0)$ a line E_1E_2 parallel and equal to OP'. Then E_2 is external to $\mathcal{R}$. For since the y coordinate of E_2 is y', and the point $(-x', y')$ lies outside or on the boundary of the domain $\mathcal{R}$ in the second quadrant, and $f(|x|, |y|)$ is an increasing function of $|x|$, it suffices to show that

$$-cf(1, 1)+x' < -x',$$

i.e.

$$2x' < cf(1, 1).$$

Since $x' \leqslant a$, this follows from (2.11). Also E_2 lies below the line $y = cf(1, 1)$, since $y' < cf(1, 1)$.

Let, now, E_1E_2 move parallel to itself, E_1 moving along the second quadrant boundary of $\mathcal{R}$. We arrive at a position, say $E_1'E_2'$, where E_2' is in the second quadrant and either lies on the line $y = cf(1, 1)$, or on the boundary of $\mathcal{R}$ as desired. The first alternative does not occur. For if it does, let E_1' be $(-\xi, \eta)$. Then E_2' is $(-\xi+x', \eta+y')$, where

$$\xi \geqslant x', \quad \eta+y' = cf(1, 1).$$

Then, since P' lies outside the domain $\mathcal{R}$,

$$f(\xi, \eta) = cf(1, 1) \leqslant f(x', y').$$

Hence $\eta \leqslant y'$, and so $2y' \geqslant cf(1, 1)$. Since the greatest possible value of y' is a, this is false from (2.11).

Suppose next that $f(1, 0) = 0$. Then the line $y = cf(1, 1)$ now plays no part in the proof, which is obvious, since the length of a line $Q'R'$ parallel to OP', with Q', R' on the second quadrant boundary of $\mathcal{R}$, varies continuously from zero to infinity.

Let Λ' be the line defined by

$$-Xy' + Yx' = 2, \tag{5.1}$$

where $P'(x', y')$ lies in the region $\mathcal{R}_1$ or $\mathcal{R}_2$.

LEMMA 11. *The line Λ' meets the axes of X, Y at points M', N' such that their distance $M'N' > OP'$.*

For
$$OM' = -2/y', \quad ON' = 2/x',$$
$$M'N' = 2(x'^2 + y'^2)^{\frac{1}{2}}/x'y' = 2OP'/x'y'. \tag{5.2}$$

Also
$$x'y' \leqslant \left(\frac{x'+y'}{2}\right)^2 \leqslant \frac{4}{(a+b)^2} < 2,$$

since the greatest value of $x'+y'$ in the regions $\mathcal{R}_1$, $\mathcal{R}_2$ occurs at the vertex $\left(\frac{2}{a+b}, \frac{2}{a+b}\right)$. Also $(a+b)^2 > a^2+b^2 = 2$. Hence the result.

LEMMA 12. *For every point of $\mathcal{R}_1$, $M'N' > 2OP'$.*

For the points of $\mathcal{R}_1$, $x'+y' < 2$, so that $x'y' < 1$, and the result is obvious from (5.2).

LEMMA 13. *If for a point $P'(x', y')$ of $\mathcal{R}_2$ the line Λ' meets the second quadrant boundary of $\mathcal{R}$ in two points, then Λ' has a length greater than $2OP'$ intercepted either between the X axis and the point of intersection of Λ' with the first quadrant boundary of $\mathcal{R}$, or between the Y axis and the point of intersection of Λ' with the third quadrant boundary of $\mathcal{R}$.*

It suffices to prove that, when $y' \leqslant x'$, the first alternative occurs, and that the intercepted length is greater than $4\sqrt{2}/(a+b)$, which is the greatest value of $2OP'$ for points of $\mathcal{R}_2$ and occurs when $x' = y' = 2/(a+b)$.

Consider, first, all the points P' of $\mathcal{R}_2$ lying on a given line through O. Since the ordinate of a point on the first quadrant boundary of $\mathcal{R}$ increases as the point moves nearer to the y axis, the intercept is obviously least when P' is at the intersection of the given line and the line $aX+bY = 2$ joining (a, b) and $\left(2/(a+b),\ 2/(a+b)\right)$.

But, for such points P' lying on $aX+bY=2$, the lines

$$-Xy'+Yx'=2$$

all pass through the point $(-b, a)$, say O'.

Let such a line, which is of course parallel to OP', meet the X axis in X', and the first quadrant boundary of $\mathfrak{R}$ in B'. Thus if P' were at (a, b), the line is $-bX+aY=2$, B' is the point $(a-b, a+b)$, say B°, and X' is the point $(-2/b, 0)$. When P' is $\left(2/(a+b), 2/(a+b)\right)$, the line is $-X+Y=a+b$ and X' is the point $\left(-(a+b), 0\right)$, which is obviously to the right of the point $(-2/b, 0)$, since $ab+b^2<a^2+b^2=2$.

Now $X'B'=X'O'+O'B'$. Clearly $X'O'$ is not less than the distance between the points O' and $(-a-b, 0)$, and so

$$X'O' \geqslant a\sqrt{2}.$$

We prove now that the least value of $O'B'$ occurs when B' is at B°, and so

$$O'B' \geqslant \sqrt{2}.$$

Let $B'B^\circ$ be produced to T, say. Then

$$0 < \text{angle } O'B^\circ T < \tfrac{1}{2}\pi.$$

For the line joining B° to the point (a, b) is perpendicular to $O'B^\circ$. Also from the convexity of the region external to $\mathfrak{R}$, B' must lie to the right of the line joining B° and (a, b), and so $O'B'>O'B^\circ$. Hence

$$X'B' \geqslant \sqrt{2}+a\sqrt{2}.$$

It remains to show that

$$\sqrt{2}+a\sqrt{2}>4\sqrt{2}/(a+b), \tag{5.3}$$

i.e.

$$(1+a)(a+b)>4.$$

We have $a^2+b^2=2$, and since $a>b>\frac{1}{2}a$ by (2.7), and so $b>\frac{1}{2}$, $1<a<\frac{1}{2}\sqrt{7}$, we have to prove that

$$(1+a)\{a+\sqrt{(2-a^2)}\}>4,$$

$$(1+a)^2(2-a^2)>(4-a-a^2)^2,$$

$$-a^4-2a^3+a^2+4a+2>a^4+2a^3-7a^2-8a+16,$$

$$a^4+2a^3-4a^2-6a+7<0,$$

i.e.

$$(a-1)(a^3+3a^2-a-7)<0.$$

The last factor is negative, since it is less than

$$\tfrac{7}{4}a+\tfrac{21}{4}-a-7=\tfrac{3}{4}a-\tfrac{7}{4}<0.$$

This proves (5.3).

The lattice L can be generated by vector addition from OP' and OP'', where P'' is some point lying in the line Λ'. For the area of the triangle $OP'P''$ is $1=\frac{1}{2}\Delta$, since Λ' is distant $2/OP'$ from OP'. Also any other point P''' on Λ' such that $P''P'''$ is an integral multiple of OP', together with P', generates L. Further, every segment of Λ' of length OP' contains a point with the generating property of P'', and not more than one such point, unless both endpoints of the segment are points of L. Next, every segment of Λ' of length $2OP'$ must contain two points, each with the generating property of P'', and three only when the end points of the segment are points of L. Hence from Lemmas 12 and 13, certainly two points of L lie on the intercepted part of Λ' referred to in Lemma 13 when Λ' meets the second quadrant boundary of $\mathfrak{R}$ in two points. When $M'N'$ lies in the interior of $\mathfrak{R}$, then by Lemmas 11 and 12, the intercepted part of Λ' contains a point of L which is an inner point of $\mathfrak{R}$, and so this case can be omitted.

Let OP' meet the first quadrant boundary of $\mathfrak{R}$ at $P(x, y)$. Let Λ, Q, R have the same relation to P that Λ', Q', R' have to P' in Lemma 10. Then by hypothesis, the area of the parallelogram $OPQR$ is not less than 2, i.e. the perpendicular from O on QR is not less than $2/OP$. Hence the line Λ', whose perpendicular distance from O is $2/OP' \leqslant 2/OP$, lies between the origin and QR. The part of Λ' in the second quadrant intercepted by the boundary of $\mathfrak{R}$ is less than or equal to OP', since it does not exceed the similar intercept on Λ, whose length is not greater than $QR=OP$, and OP does not exceed OP'. Hence there cannot be two points of L in the Λ' intercept by $\mathfrak{R}$. Hence, unless equality occurs above, i.e. when P' coincides with P, and the area $OPQR$ is 2, i.e. when P is at the points (b, a) or (a, b), the line Λ' contains a point of L other than O which is an inner point of $\mathfrak{R}$. This completes the proof of Theorem 1.

We have next to see if $\mathfrak{R}$ is such a region that

$$\text{area parall. } OPQR \geqslant 2, \tag{5.4}$$

when P lies in the boundary arc of $\mathfrak{R}$ joining (b, a), (a, b). The midpoint of the arc is (c, c), and by symmetry it suffices to consider only the arc joining (b, a), (c, c).

The condition that the area $OPQR$ is not less than 2 in the particular case when P is the point (c, c) can be written as

$$f\left(\frac{1}{c^2}-\tfrac{1}{2},\ \frac{1}{c^2}+\tfrac{1}{2}\right) \leqslant f(1, 1). \tag{5.5}$$

For it suffices if the line $-x+y=2/c$ cuts off a chord in the second quadrant not exceeding $c\sqrt{2}$ in length. The line meets the boundary of $\mathcal{R}$ at the points in the second quadrant

$$\left(-\frac{1}{c}+\zeta,\ \frac{1}{c}+\zeta\right), \quad \left(-\frac{1}{c}-\zeta,\ \frac{1}{c}-\zeta\right),$$

where ζ is given by

$$g(\zeta) \equiv f\left(\frac{1}{c}-\zeta,\ \frac{1}{c}+\zeta\right) = cf(1, 1),$$

and must satisfy $0 < 2\zeta \leqslant c$ from the length of the chord.

When $\zeta = 0$, $g(0) = f\left(\frac{1}{c}, \frac{1}{c}\right) = \frac{1}{c} f(1, 1) > cf(1, 1)$, since $c < 1$. Since $g(\zeta)$ is a steadily decreasing function of ζ, clearly $\zeta \leqslant \frac{1}{2}c$ if $g(\frac{1}{2}c) \leqslant cf(1, 1)$. This gives the result (5.5). It can be put in another form by writing $a = cm$, $b = cn$, so that $c^2(m^2+n^2) = 2$. Let $m > 0$, $n > 0$ satisfy

$$f(m, n) = f(m-n,\ m+n) = f(1, 1). \tag{5.6}$$

Then the condition becomes

$$f\left(\tfrac{1}{2}(m^2+n^2-1),\ \tfrac{1}{2}(m^2+n^2+1)\right) \leqslant f(1, 1). \tag{5.7}$$

Hence if a function $f(x, y)$ which satisfies I, II, III of § 2 also satisfies (5.6), (5.7), it is suggested that (3.4) with $\Delta = 2$ sometimes gives the best possible result.

It is clear that Theorem 2 follows from the proof of Theorem 1.

It does not appear, however, to be an easy matter to find fairly general conditions under which (3.14) is satisfied, i.e. for a best possible result. A little thought shows that this is not surprising. Thus suppose that the region $\mathcal{R}$ is finite. All the results proved remain unchanged if $\mathcal{R}$ is altered by a suitable long narrow extension along the axes of X, Y. Then since a, b are unaltered, there is the possibility, e.g. with (3.12), of choosing ξ, η so that Y is small enough for the corresponding lattice point to be in the new $\mathcal{R}$.

Write $$X = a\xi + b\eta, \quad Y = b\xi - a\eta. \tag{5.8}$$

Then a result that is sometimes useful is given in

LEMMA 14. *If $\mathcal{R}$ is finite, then* (3.12), (3.13) *are critical lattices when for some integer $D > 0$,*

$$f\{\sqrt{(2D+1)}, 0\} \geqslant cf(1, 1), \tag{5.9}$$

and for the finite number of lattice points given by (5.8), *where*

$$0 < \xi^2 + \eta^2 \leqslant D, \tag{5.10}$$

$$f(|X|, |Y|) \geqslant cf(1, 1). \tag{5.11}$$

Three particular sets of conditions will be deduced by taking $D = 4$, 12, 32 respectively, namely:

I. $$f(3, 0) \geqslant cf(1, 1); \tag{5.12}$$

II. $$f(5, 0) \geqslant cf(1, 1), \tag{5.13}$$

and $$f(2a+b, 2b-a) \geqslant cf(1, 1), \tag{5.14}$$

and (5.14) *is certainly satisfied if $j = b/a \geqslant \frac{2}{3}$*;

III. $$f(8, 0) \geqslant cf(1, 1), \tag{5.15}$$

$$j \geqslant \tfrac{2}{3}, \tag{5.16}$$

$$f(3b-2a, 2b+3a) \geqslant cf(1, 1), \tag{5.17}$$

$$f(|4b-3a|, 3b+4a) \geqslant cf(1, 1). \tag{5.18}$$

Proof of Lemma 14. By (5.10) we need consider only the case when $\xi^2+\eta^2 > D$, or, since

$$X^2 + Y^2 = (a^2+b^2)(\xi^2+\eta^2) = 2(\xi^2+\eta^2),$$

the case when $X^2+Y^2 \geqslant 2D+2$. Since $c < 1$, (5.11) is clearly satisfied when both $|X| \geqslant 1$ and $|Y| \geqslant 1$. Hence we may suppose that, say, $|Y| < 1$ and then $X^2 > 2D+1$, and (5.9) shows that (5.11) holds again.

Write now $\xi^2+\eta^2 = d$, and suppose that $d \leqslant D$.

Next, I follows on taking $D = 4$. For if $d = 1, 2, 4$, the resulting points (5.8) are either boundary points such as (a, b), $(a+b, -a+b)$, etc., or obvious external points such as $(2a, 2b)$, etc.

For II, take $D = 12$. We need only consider $d = 5, 8, 9, 10$. For $d = 5$, we need only consider $(2a-b, a+2b)$, $(2a+b, 2b-a)$. The first is an external point since $2a-b > a-b$, $a+2b > a+b$. The second is either an external point or a boundary point by hypothesis. A sufficient condition for this is $2b-a \geqq a-b$ or $3b \geqq 2a$. The values $d = 8, 9$ lead to obvious external points; and $d = 10$ also, since $(3b-a, 3a+b)$ is one since $3b-a > a-b$, i.e. $b > \frac{1}{2}a$, while $(3b+a, 3a-b)$ is obviously an external point.

For III, take $D = 32$. We need consider now only $d = 13, 16, 17, 18, 20, 25, 26, 29, 32$. When $d = 13$ it suffices to remark that $(3b-2a, 2b+3a)$ is an external point from (5.17), while $(3a-2b, 3b+2a)$ is obviously one. The values $d = 16, 18, 32$ lead to obvious external points; also $d = 17, 26$ since $b > 2a$ and so $4b-a > a-b$, $5b-a > a-b$; also $d = 20$ since $b/a > \frac{2}{3}$ and so $4b-2a > a-b$; $d = 25$ since $\xi = 5$, $\eta = 0$ leads to an external point, as does $\xi = 3$, $\eta = 4$ while $\xi = 4$, $\eta = 3$, comes under (5.18); and finally $d = 29$ since $2b > a$ and so $5b-2a > a-b$.

Suppose next that $\mathfrak{R}$ is infinite. As seen later in the paper, it is possible in some cases to verify directly that the conditions (3.14) or (3.16) are satisfied.

An interesting case arises when b/a is rational. Then an infinity of points of L lie on the coordinate axes, and the inequality (3.4) is satisfied trivially. Even then, as seen later, we may still have a best possible result in the sense that if the constant on the right-hand side of (3.4) is diminished, the lattices (3.12), (3.13) have no points in the new region so defined other than the points of L on the coordinate axes.

6. *The circular quadrilateral.*

THEOREM 3. *Let*

$$f(x, y) = x+y+[(\lambda^2-1)(x^2+y^2)+2xy]^{\frac{1}{2}}, \tag{6.1}$$

where $1 \leqslant \lambda < \sqrt{2}$. *Then a point other than the origin of every lattice of determinant* $\Delta > 0$ *lies in the region*

$$f(|x|, |y|) \leqslant [\lambda^2+2+(-\lambda^4+12\lambda^2-4)^{\frac{1}{2}}]^{\frac{1}{2}}\Delta^{\frac{1}{2}}. \tag{6.2}$$

This is the best possible result. There are exactly two critical lattices.

[When we write $\lambda^2-1 = \mu$ in (6.1), where $0 \leqslant \mu \leqslant 1$, the term before $\Delta^{\frac{1}{2}}$ in (6.2) becomes

$$[\mu+3+(-\mu^2+10\mu+7)^{\frac{1}{2}}]^{\frac{1}{2}}.]$$

Expressed geometrically, this result means:

Let there be a family of circular quadrilaterals whose centres are at the origin, whose vertices are on the axes, and whose sides are arcs of equal circles which all have their convex parts away from the origin and which cut the axes at an angle $\sec^{-1}\lambda$, *where clearly* $1 \leqslant \lambda < \sqrt{2}$. *Then every lattice of determinant* Δ *has one of its points other than O in the quadrilateral, with the centre of one of its circular arcs at* $(g\sqrt{\frac{1}{2}\Delta}, g\sqrt{\frac{1}{2}\Delta})$, *where*

$$g^2 = \frac{2\lambda^2+4+2(-\lambda^4+12\lambda^2-4)^{\frac{1}{2}}}{(2-\lambda^2)^2}. \tag{6.3}$$

This is the best possible result.

The equation of the first quadrant boundary of the quadrilateral can be written as

$$(x-g)^2+(y-g)^2 = \lambda^2 g^2, \tag{6.4}$$

where g is a parameter. Hence

$$g^2(2-\lambda^2)-2g(x+y)+x^2+y^2 = 0,$$

$$(2-\lambda^2)g = x+y \pm [(x+y)^2-(2-\lambda^2)(x^2+y^2)]^{\frac{1}{2}}.$$

On considering the intersections with the x-axis, it is clear that the $+$ sign must be taken and so

$$(2-\lambda^2)g = x+y+[(\lambda^2-1)(x^2+y^2)+2xy]^{\frac{1}{2}}.$$

We take then

$$f(x, y) = x+y+[(\lambda^2-1)(x^2+y^2)+2xy]^{\frac{1}{2}},$$

and it is easy to see that all the conditions I, II, III of § 2 are satisfied, e.g., $f(1-y, 1+y)$ is a decreasing function of y in $0 \leqslant y \leqslant 1$, since $(\lambda^2-1)(1+y^2)+1-y^2$ is so from $\lambda^2 < 2$.

Since the circle (6.4) passes through the points (a, b), $(a+b, a-b)$, (c, c), the equations giving a, b, c can be written as

$$\left.\begin{aligned}(a-g)^2+(b-g)^2 &= \lambda^2 g^2,\\ (a+b-g)^2+(a-b-g)^2 &= \lambda^2 g^2,\end{aligned}\right\} \tag{6.5}$$

$$2(c-g)^2 = \lambda^2 g^2, \tag{6.6}$$

$$a^2+b^2 = 2. \tag{6.7}$$

Hence from (6.5)

$$\left.\begin{aligned} 2-2g(a+b) &= (\lambda^2-2)g^2, \\ 4-4ga &= (\lambda^2-2)\,g^2. \end{aligned}\right\} \tag{6.8}$$

Clearly
$$-4gb = (\lambda^2-2)\,g^2, \tag{6.9}$$

and so
$$32g^2 = [4-(\lambda^2-2)\,g^2]^2+[(\lambda^2-2)\,g^2]^2,$$

i.e.
$$(\lambda^2-2)^2g^4-4(\lambda^2+2)g^2+8 = 0,$$

and
$$(-\lambda^2+2)^2g^2 = 2\lambda^2+4\pm(-4\lambda^4+48\lambda^2-16)^{\frac{1}{2}}. \tag{6.10}$$

We show that the + sign must be taken. For, from (2.7), $b/a > \frac{1}{2}$, and so from (6.8), (6.9),

$$\frac{(2-\lambda^2)\,g^2}{4+(2-\lambda^2)\,g^2} > \tfrac{1}{2},$$

i.e.
$$(2-\lambda^2)\,g^2 > 4.$$

If we take the − sign in (6.10), this gives

$$2\lambda^2+4-(-4\lambda^4+48\lambda^2-16)^{\frac{1}{2}} > 4(2-\lambda^2),$$

i.e.
$$(3\lambda^2-2)^2 > (-\lambda^4+12\lambda^2-4),$$

i.e.
$$(\lambda^2-2)(10\lambda^2-4) > 0.$$

This, however, does not hold for $1 \leqslant \lambda^2 \leqslant 2$.

Let now $P(x, y)$ be any point of the first quadrant boundary lying between (b, a), (c, c); and let QR be the line in the second quadrant equal and parallel to OP and whose ends Q, R lie on the boundary of the region. Since the equation of OP is $-yX+xY = 0$, and the centre of the circular boundary through Q, R is at $(-g, g)$, the perpendicular from O on QR is equal to

$$\frac{gx+gy}{(x^2+y^2)^{\frac{1}{2}}} - (\lambda^2g^2-\tfrac{1}{4}OP^2)^{\frac{1}{2}}.$$

Hence the area of the parallelogram $OPQR$ is

$$g(x+y)-OP(\lambda^2g^2-\tfrac{1}{4}OP^2)^{\frac{1}{2}}.$$

Since
$$2g(x+y) = x^2+y^2+(2-\lambda^2)\,g^2,$$

the area, on writing $OP = r$, becomes

$$\tfrac{1}{2}[r^2+(2-\lambda^2)g^2]-r(\lambda^2g^2-\tfrac{1}{4}r^2)^{\frac{1}{2}}.$$

Hence we wish to prove that, say,

$$E = r^2+(2-\lambda^2)g^2-r(4\lambda^2g^2-r^2)^{\frac{1}{2}} \geqslant 4$$

for $2c^2 \leqslant r^2 \leqslant 2$. We know from our theory that equality is attained when $r^2 = 2$, i.e.,

$$(2-\lambda^2)g^2-2(2\lambda^2g^2-1)^{\frac{1}{2}} = 2.$$

It suffices to show that for $c\sqrt{2} \leqslant r \leqslant \sqrt{2}$, $dE/dr < 0$, *i.e.*,

$$2r-(4\lambda^2g^2-r^2)^{\frac{1}{2}}+r^2(4\lambda^2g^2-r^2)^{-\frac{1}{2}} < 0,$$

i.e.

$$2r(4\lambda^2g^2-r^2)^{\frac{1}{2}}+2r^2-4\lambda^2g^2 < 0,$$

$$r^2(4\lambda^2g^2-r^2) < (2\lambda^2g^2-r^2)^2$$

(provided that $2\lambda^2g^2 \geqslant r^2$),

i.e.

$$r^4-4\lambda^2g^2r^2+2\lambda^4g^4 > 0,$$

i.e.

$$(r^2-2\lambda^2g^2)^2-2\lambda^4g^4 > 0.$$

This will certainly be satisfied if $r^2/\lambda^2g^2 < 2-\sqrt{2}$, i.e.,

$$\lambda^2g^2 > r^2\left(\frac{2+\sqrt{2}}{2}\right) \geqslant 2+\sqrt{2}.$$

Then also $2\lambda^2g^2 \geqslant 2 \geqslant r^2$, as required previously.

Hence it suffices if

$$\lambda^2\left[\frac{2\lambda^2+4+2\sqrt{(-\lambda^4+12\lambda^2-4)}}{(2-\lambda^2)^2}\right] > 2+\sqrt{2}$$

when $1 \leqslant \lambda^2 < 2$. This is obvious since the left-hand side is greater than $\lambda^2(2\lambda^2+4) \geqslant 6$.

We show finally that the lattices (3.12), (3.13) are critical lattices by applying II of Lemma 14.

We show first that $j > \frac{2}{3}$. By (6.8), (6.9) this becomes

$$(2-\lambda^2)g^2/[(2-\lambda^2)g^2+4] > \tfrac{2}{3},$$

i.e.

$$(2-\lambda^2)g^2 > 8.$$

From (6.10), this is so if

$$2\lambda^2+4+(-4\lambda^4+48\lambda^2-16)^{\frac{1}{2}} > 8(2-\lambda^2),$$

i.e.

$$(-4\lambda^4+48\lambda^2-16)^{\frac{1}{2}} > 12-10\lambda^2. \qquad (6.11)$$

This is obvious if $\lambda^2 \geqslant 1{\cdot}2$, so that we may suppose $1 \leqslant \lambda^2 < 1{\cdot}2$ But (6.11) also holds then, for it becomes

$$-\lambda^4+12\lambda^2-4 > 36-60\lambda^2+25\lambda^4,$$

i.e.

$$0 > 40-72\lambda^2+26\lambda^4,$$

$$0 > 40(1-\lambda^2)+2\lambda^2(-16+13\lambda^2),$$

and this is obviously true since $13\lambda^2 < 16$.

We have now to verify (5.13), which becomes here

$$5\{1+\sqrt{(\lambda^2-1)}\} \geqslant c(2+\lambda\sqrt{2}) = g(2-\lambda^2),$$

as is obvious from a figure, or also from (6.6). Hence, from (6.10), we have to show that

$$5\{1+\sqrt{(\lambda^2-1)}\} \geqslant (2\lambda^2+4+\sqrt{[-4\lambda^4+48\lambda^2-16]})^{\frac{1}{2}}.$$

The square of the right-hand side is

$$2\lambda^2+4+\{-4(6-\lambda^2)^2+128\}^{\frac{1}{2}} \leqslant 4+4+12 = 20 < 25;$$

whence the result.

7. *The star-shaped octagon.*

I now prove

THEOREM 4. *Let*

$$f(x, y) = \min [x+my, y+mx], \tag{7.1}$$

where the parameter m satisfies $1/\sqrt{3} \leqslant m \leqslant 1$, so that

$$\left.\begin{aligned} f(x, y) &= x+my \quad \textit{when} \quad 0 \leqslant x \leqslant y, \\ f(x, y) &= mx+y \quad \textit{when} \quad 0 \leqslant y \leqslant x. \end{aligned}\right\} \tag{7.2}$$

Then a point other than the origin of every lattice of determinant $\Delta > 0$ lies in the region

$$f(|x|, |y|) \leqslant \frac{1+2m-m^2}{(5-4m+m^2)^{\frac{1}{2}}}\,\Delta^{\frac{1}{2}}. \tag{7.3}$$

The right-hand constant is the best possible.

Expressed geometrically this means:

Let there be a family of star-shaped equal-sided octagons with centres at the origin, remoter and nearer vertices on the coordinate axes and the lines

$y = \pm x$ respectively, and with given vertical angle $2\theta = 2\tan^{-1} m$, so that $60° \leqslant 2\theta \leqslant 90°$. Then every lattice of determinant $\Delta > 0$ has one of its points other than O lying in the star of the family with a near vertex at (c, c) and a far vertex at $\left(0,\ c\left(1+\frac{1}{m}\right)\right)$, where

$$(1+m)c = \frac{1+2m-m^2}{(5-4m+m^2)^{\frac{1}{2}}}\Delta^{\frac{1}{2}}.$$

This is the best possible result. There are two critical lattices except in the special case when $m = 1/\sqrt{3}$, when there are four.

It is easily seen that $f(x, y)$ is a homogeneous function of dimension $p = 1$ satisfying the conditions of § 2. It suffices to prove that $f(1-y, 1+y)$ decreases steadily for $0 < y < 1$. This is clear from

$$f(1-y,\ 1+y) = 1-y+m(1+y) = 1+m-(1-m)y.$$

The equations defining a, b, c become

$$m(a+b)+(a-b) = c+mc, \quad ma+b = c+mc,$$

$$a^2+b^2 = 2.$$

From the first two

$$a-(2-m)b = 0.$$

Hence

$$b^2 = 2/(m^2-4m+5), \quad a^2 = 2(2-m)^2/(m^2-4m+5), \tag{7.4}$$

$$c^2 = 2(1+2m-m^2)^2/(m^2-4m+5)(1+m)^2. \tag{7.5}$$

As in the proof of Theorem 1, let $P(x, y)$ be any point of the first quadrant boundary $x+my = c(1+m)$ lying between (b, a) and (c, c). Let $(-q, r)$ be the middle point of the chord in the second quadrant which is equal and parallel to OP. Then, as in (3.6), (3.7), the coordinates of the end of the chord are

$$(-q+\tfrac{1}{2}x,\ r+\tfrac{1}{2}y), \quad (-q-\tfrac{1}{2}x,\ r-\tfrac{1}{2}y),$$

where

$$(q-\tfrac{1}{2}x)+m(r+\tfrac{1}{2}y) = c(1+m),$$

$$m(q+\tfrac{1}{2}x)+(r-\tfrac{1}{2}y) = c(1+m).$$

We have to investigate conditions such that, for all the points P above,

$$rx+qy \geqslant 2.$$

Clearly

$$q+mr = \tfrac{1}{2}(x-my)+c(1+m),$$

$$mq+r = \tfrac{1}{2}(-mx+y)+c(1+m).$$

Multiply by $y-mx$, $x-my$ respectively and add. Then

$$(1-m^2)(rx+qy) = (x-my)(y-mx)+c(x+y)(1-m^2).$$

We require then the condition that

$$(x-my)(y-mx)+c(x+y)(1-m^2) \geqslant 2(1-m^2). \tag{7.6}$$

This must hold for $(x, y) = (c, c)$, and so

$$\frac{1-m}{1+m}\,c^2+2c^2 \geqslant 2,$$

i.e. $$c^2 \geqslant (2+2m)/(3+m).$$

Hence, from (7.5),

$$\frac{(1+2m-m^2)^2}{(1+m)^2(5-4m+m^2)} \geqslant \frac{1+m}{3+m},$$

i.e. $$(1+2m-m^2)^2(3+m) \geqslant (1+m)^3(5-4m+m^2).$$

This reduces to

$$6m^3-6m^2-2m+2 \leqslant 0,$$

$$(1-m)(1-3m^2) \leqslant 0, \tag{7.7}$$

i.e. $1 \geqslant m \geqslant 1/\sqrt{3}$, *i.e.* $45^\circ \geqslant \theta \geqslant 30^\circ$.

The value $m = 1/\sqrt{3}$ leads to

$$c^2 = (2+2\sqrt{3})/(3\sqrt{3}+1) = (8+2\sqrt{3})/13 = \cdot 8818\ldots,$$

$$c = \cdot 93\ldots. \tag{7.8}$$

Write now $x = c-mz$, $y = c+z$, so that as P goes from (b, a) to (c, c), z goes from, say, z_0 to 0, where $c-mz_0 = b$ and so $z_0 > 0$. Then for these values of z, we require from (7.6) that

$$[c(1-m)-2zm][c(1-m)+z(1+m^2)]+c[2c+z(1-m)](1-m^2) \geqslant 2(1-m^2),$$

i.e.

$$0 \geqslant 2m(1+m^2)z^2+zc[2m(1-m)-(1-m)(1+m^2)-(1-m)(1-m^2)]$$
$$+2(1-m^2)-c^2(1-m)^2-2c^2(1-m^2).$$

This inequality, since it holds for $z=0$ and becomes an equality for $z=z_0$, takes the form, say,

$$0 \geqslant (z-z_0)(z+z_1),$$

where $z_1 \geqslant 0$. Hence it is satisfied for $0 \leqslant z \leqslant z_0$.

Finally, it is easily seen that the condition (5.12) for a critical lattice is satisfied. For it becomes $3m \geqslant c(1+m)$, and since $m \geqslant 1/\sqrt{3}$, $3m \geqslant 1+m$.

Hence we have found the best possible result for a star-shaped equal-sided octagon of acute angle lying between 60° and 90°.

When the angle is 60°, *i.e.* $3m^2=1$, there are two additional critical lattices as given in (3.15). The condition for these is, by (3.16), that

$$\min\,[m|X|+|Y|;\ |X|+m|Y|] \geqslant 1+m, \tag{7.9}$$

where $$X=\xi+\left(\tfrac{1}{2}+\frac{1}{c^2}\right)\eta, \qquad Y=\xi+\left(\tfrac{1}{2}-\frac{1}{c^2}\right)\eta,$$

and ξ, η run through all integer values except 0, 0. Write $\zeta=\xi+\frac{1}{2}\eta$; then

$$X=\zeta+\frac{\eta}{c^2}, \qquad Y=\zeta-\frac{\eta}{c^2}.$$

The condition (7.9) becomes

$$\min\left[m\left|\zeta+\frac{\eta}{c^2}\right|+\left|\zeta-\frac{\eta}{c^2}\right|,\quad \left|\zeta+\frac{\eta}{c^2}\right|+m\left|\zeta-\frac{\eta}{c^2}\right|\right] \geqslant 1+m.$$

We may clearly suppose that $\eta \geqslant 0$, and then also that $\zeta \geqslant 0$ on replacing ξ, if need be, by $-\xi-\eta$. Then (7.9) becomes

$$m\left|\zeta+\frac{\eta}{c^2}\right|+\left|\zeta-\frac{\eta}{c^2}\right| \geqslant 1+m. \tag{7.10}$$

This is true if $\eta=0$, since then $\zeta \geqslant 1$. Suppose first that $\zeta \geqslant \eta/c^2$; then (7.10) is certainly satisfied if

$$2m\eta/c^2 \geqslant 1+m,$$

i.e. $$\eta \geqslant (1+m)c^2/2m = 1{\cdot}20\ldots > 1.$$

Hence we need only consider $\eta = 1$, $\zeta \geqslant 1/c^2$, *i.e.* $\zeta \geqslant \frac{3}{2}$. Then condition (7.10) is satisfied if

$$m\left(\tfrac{3}{2}+\frac{1}{c^2}\right)+\tfrac{3}{2}-\frac{1}{c^2} \geqslant 1+m,$$

i.e. $$\tfrac{1}{2}(1+m) \geqslant (1-m)/c^2,$$

i.e.

$$c^2 \geqslant 2(1-m)/(1+m) = 2(\sqrt{3}-1)/(\sqrt{3}+1) = (\sqrt{3}-1)^2,$$

i.e. $$c \geqslant \sqrt{3}-1 = \cdot 732\ldots,$$

which is satisfied by (7.8).

Suppose next that $\zeta < \eta/c^2$. The condition (7.10) is certainly satisfied if

$$2m\zeta \geqslant 1+m,$$

i.e. $$\zeta \geqslant (1+m)/2m = 1\cdot 3\ldots.$$

Hence we need only consider $\zeta = \frac{1}{2}, \eta = 1, 3, 5, \ldots$, and $\zeta = 1, \eta = 2, 4, 6, \ldots$.

When $\zeta = \frac{1}{2}$, condition (7.10) is

$$\tfrac{1}{2}(m-1)+\eta(m+1)/c^2 \geqslant 1+m.$$

This holds since $\frac{1}{2}(m-1)+(m+1)/c^2 = 1+m$.

When $\zeta = 1$, condition (7.10) is

$$m-1+(m+1)\eta/c^2 \geqslant 1+m,$$

and this holds since

$$m-1+2(m+1)/c^2 \geqslant 1+m,$$

which follows from $c^2 < 1$.

This completes the proof.

8. *The case* $f(x, y) = x^p+y^p$.

This, when $0 < p < 1$, is the non-convex case not considered by Minkowski. It is easily verified that $f(x, y)$ satisfies the conditions of § 2. Thus, for $0 \leqslant y \leqslant 1$, $(1+y)^p+(1-y)^p$ is a steadily decreasing function of y; also the region $x^p+y^p \geqslant 1$, $x \geqslant 0$, $y \geqslant 0$ is convex.

The equations determining a, b, c are

$$(a+b)^p+(a-b)^p = a^p+b^p = 2c^p, \tag{8.1}$$

$$a^2+b^2 = 2, \quad a > b > 0, \tag{8.2}$$

and so, if $j = b/a < 1$,

$$(1+j)^p+(1-j)^p = 1+j^p. \tag{8.3}$$

I prove now

THEOREM 5. *Every lattice of determinant $\Delta > 0$ has one of its points other than O in the region $\mathfrak{R}$,*

$$|x|^p+|y|^p \leqslant 2^{1-\frac{1}{2}p}\,\Delta^{\frac{1}{2}p}\,c^p, \tag{8.4}$$

when $0 \leqslant p \leqslant \frac{1}{3}$ or $p = \frac{1}{2}$.

The constant on the right-hand side of (8.4) is best possible when $p = \frac{1}{2}$ or $p = \frac{1}{3}$.

Presumably (8.4) holds for all p in $0 < p < 1$, and the method developed in this paper seems to be adequate in principle for the proof. It requires for $0 < p < 1$ the apparently very simple conjecture*

LEMMA 15. *Let a curve be given by $F(U, V) = 0$, where*

$$F(U, V) = [(U+1)^p-1][(V+1)^p-1]-[1-U^p][1-V^p]. \tag{8.5}$$

Then every line

$$U+V = 2d \tag{8.6}$$

meets the curve in at most two points lying in the first quadrant.

I prove this later for $p = \frac{1}{2}$.

For Theorem 5, however, rather less is required than this lemma. It will suffice to prove it for the special case when

$$2d = j+(1/j)-1, \tag{8.7}$$

and this I do later for $0 < p \leqslant \frac{1}{3}$.

We now carry out the general procedure of § 3 leading to equations (3.5) to (3.9). Denote by A, B the points (b, a), (a, b). Let $P(x, y)$, where

$$x^p+y^p = 2c^p, \tag{8.8}$$

be any point in the arc AB of the boundary of $\mathfrak{R}$; and let $(-q, r)$ be the middle point of the line QR which is equal and parallel to OP and has its

* As already remarked, this has been since proved by Watson (**6**). See the addendum to § 8. [(8.4) is the best possible result for $1 > p > \cdot 3295\ldots$.]

ends Q, R on the second quadrant boundary of $\mathcal{R}$. Then

$$q > \tfrac{1}{2}x, \quad r > \tfrac{1}{2}y. \tag{8.9}$$

$$(q-\tfrac{1}{2}x)^p + (r+\tfrac{1}{2}y)^p = 2c^p. \tag{8.10}$$

$$(q+\tfrac{1}{2}x)^p + (r-\tfrac{1}{2}y)^p = 2c^p. \tag{8.11}$$

We have to prove that

$$qy + rx \geqslant 2, \tag{8.12}$$

equality occurring only at the points A, B. Put

$$q = x(U+\tfrac{1}{2}), \quad r = y(V+\tfrac{1}{2}), \tag{8.13}$$

then
$$U \geqslant 0, \quad V \geqslant 0,$$

$$x^p U^p + y^p (V+1)^p = 2c^p, \tag{8.14}$$

$$x^p (U+1)^p + y^p V^p = 2c^p, \tag{8.15}$$

and we wish to prove that

$$U + V + 1 \geqslant 2/xy. \tag{8.16}$$

LEMMA 16. *U, V satisfy the equation*

$$F(U, V) = [(U+1)^p - 1][(V+1)^p - 1] - [1-U^p][1-V^p] = 0. \tag{8.17}$$

Also
$$0 \leqslant U, V \leqslant 1, \tag{8.18}$$

and as U increases from 0 to 1, V decreases steadily from 1 to 0.

From (8.14), (8.15), (8.8)

$$x^p (U^p - 1) + y^p \left((V+1)^p - 1 \right) = 0, \tag{8.19}$$

$$x^p \left((U+1)^p - 1 \right) + y^p (V^p - 1) = 0. \tag{8.20}$$

The equation $F(U, V) = 0$ follows on eliminating x^p, y^p. It shows at once that neither one of U, V, nor both, can be greater than unity.

The second part follows since differentiation shows that

$$[(U+1)^p - 1]/(1-U^p) \quad \text{and} \quad (1-V^p)/[(V+1)^p - 1]$$

are increasing and decreasing functions of U, V respectively. Thus

$$(1-U^p)(U+1)^{p-1} + U^{p-1}[(U+1)^p - 1] \geqslant 0,$$

i.e. $$(U+1)^{p-1}-U^{p-1}+U^{p-1}(U+1)^{p-1} \geqslant 0,$$

follows from $$U^{1-p}-(U+1)^{1-p}+1 \geqslant 0,$$

since the left-hand side is obviously an increasing function of U for $U \geqslant 0$.

LEMMA 17. *U, V satisfy the inequalities*

$$(1/j)-1 \leqslant U,\ V \leqslant j. \tag{8.21}$$

We first establish the inequalities

$$(U+1)^p+j^p V^p \geqslant j^p+1, \tag{8.22}$$

$$(V+1)^p+j^p U^p \geqslant j^p+1, \tag{8.23}$$

the proof of which is very simple. For since P lies on the boundary of $\mathfrak{R}$ between (b, a) and (a, b), we have

$$1/j \geqslant x/y \geqslant j, \tag{8.24}$$

the equality sign occurring only at the points (b, a) or (a, b), *i.e.* A or B. Hence from (8.19),

$$[(V+1)^p-1]/(1-U^p) = x^p/y^p \geqslant j^p. \tag{8.25}$$

This proves (8.23). Similarly for (8.22).

Denote by C, D the points $(U, V) = (1/j-1, j)$, $(j, 1/j-1)$ on the line (8.6), (8.7). It suffices now* to show that U increases steadily from $1/j-1$ to j as y/x decreases from $1/j$ to j. Write

$$(y/x)^p = z;$$

then $$1-U^p = z\big((V+1)^p-1\big),$$

$$(U+1)^p-1 = z(1-V^p),$$

Differentiating, we have

$$-pU^{p-1}\frac{dU}{dz} = (V+1)^p-1+pz(V+1)^{p-1}\frac{dV}{dz},$$

$$p(U+1)^{p-1}\frac{dU}{dz} = 1-V^p-pzV^{p-1}\frac{dV}{dz}.$$

Multiply by V^{p-1}, $(V+1)^{p-1}$ respectively and add. Then

$$p[(U+1)^{p-1}(V+1)^{p-1}-U^{p-1}V^{p-1}]\frac{dU}{dz} = (V+1)^{p-1}-V^{p-1}+V^{p-1}(V+1)^{p-1}.$$

* This proof is due to Dr. Davenport and is simpler than my original one.

As seen in Lemma 16, the right-hand side is never negative. Hence

$$\frac{dU}{dz} < 0, \quad \text{when} \quad V \neq 0.$$

Clearly from (8.20) the points C, D correspond to the points A, B.

We now modify the condition (8.16). From (8.8), it is clear that xy increases steadily from ab to c^2 as P moves from (b, a) to (c, c). Hence it suffices to show that

$$U+V+1 \geqslant 2/ab = (a^2+b^2)/ab = j+1/j, \tag{8.26}$$

which of course is rather more than (8.16).

The line

$$U+V+1 = j+1/j \tag{8.27}$$

meets the curve $F(U, V) = 0$ at the points C, D, and the axes of U, V at points $U = j+1/j-1 > 1$, $V = 0$ and $V = j+1/j-1 > 1$, $U = 0$. Hence it will suffice to prove that the line (8.27) meets the curve $F(U, V) = 0$ in only the two points C, D. I prove, more generally,

LEMMA 18. *If $p = \frac{1}{2}$, $U+V$ increases steadily as the point (U, V) moves along the curve $F(U, V) = 0$ from the point $(0, 1)$ to (t, t) where*

$$(t+1)^p+t^p = 2. \tag{8.28}$$

An alternative enunciation is that the curve $F(U, V) = 0$ meets any line $U+V = 2d$ in at most two points in the first quadrant.

Now, for $X \geqslant 0$, $Y \geqslant 0$, $0 < p < 1$,

$$\tfrac{1}{2}(X^p+Y^p) \leqslant [\tfrac{1}{2}(X+Y)]^p,$$

equality occurring only when $X = Y$. It is clear then from this that if $d < \frac{1}{2}$, $U+V = 2d$ leads to

$$(U+1)^p-1 < 1-V^p,$$

etc., and so there are no real intersections; and that if $d = \frac{1}{2}$, there are only two, namely $(U, V) = (1, 0)$, $(0, 1)$.

The curve $F(U, V) = 0$ passes through the point (t, t). Considerations of continuity show now that there are at least two real points of intersection when $\frac{1}{2} < d \leqslant t$, which we suppose hereafter.

I take now $p = \frac{1}{2}$, so that

$$[(U+1)^{\frac{1}{2}}-1][(V+1)^{\frac{1}{2}}-1] = [1-U^{\frac{1}{2}}][1-V^{\frac{1}{2}}]$$

and show that $U+V$ steadily increases as U increases from 0 to t. Here $t = \frac{9}{16}$, since now $(t+1)^{\frac{1}{4}}+t^{\frac{1}{4}} = 2$.

Put $$U^{\frac{1}{4}} = (1-r^2)/2r.$$

Then since $r^2+2rU^{\frac{1}{4}} = 1$, we can take

$$r = -U^{\frac{1}{4}}+(U+1)^{\frac{1}{4}}.$$

Hence as U increases from 0 to $\frac{9}{16}$, r decreases from 1 to $\frac{1}{2}$. Similarly if

$$V^{\frac{1}{4}} = (1-s^2)/2s, \quad s = -V^{\frac{1}{4}}+(V+1)^{\frac{1}{4}},$$

then as V decreases from 1 to $\frac{9}{16}$, s increases from $\sqrt{2}-1$ to $\frac{1}{2}$.

The U, V curve becomes

$$\left(\frac{1+r^2}{2r}-1\right)\left(\frac{1+s^2}{2s}-1\right) = \left(1-\frac{1-r^2}{2r}\right)\left(1-\frac{1-s^2}{2s}\right),$$

i.e. $$(r^2+2r-1)(s^2+2s-1) = (r-1)^2(s-1)^2,$$

and reduces to $$r^2+s^2 = 2rs(r+s).$$

We have now to show that as r decreases from 1 to $\frac{1}{2}$ and so s increases from $\sqrt{2}-1$ to $\frac{1}{2}$,

$$\left(\frac{1-r^2}{2r}\right)^2+\left(\frac{1-s^2}{2s}\right)^2,$$

or, say, $$E = r^2+s^2+\frac{1}{r^2}+\frac{1}{s^2}$$

steadily increases.

Put $$r = R\cos\theta, \quad s = R\sin\theta.$$

Then $$R\sin 2\theta(\sin\theta+\cos\theta) = 1,$$

where θ increases from $\tan^{-1}(\sqrt{2}-1)$ to $\tan^{-1}1$, i.e. from $\frac{1}{8}\pi$ to $\frac{1}{4}\pi$. Also

$$E = R^2+1/R^2\sin^2\theta\cos^2\theta.$$

Put $$X = \sin 2\theta,$$

so that X goes from $1/\sqrt{2}$ to 1. Since $1/R^2 = X^2(1+X)$,

$$E = 4(1+X)+1/X^2(1+X).$$

Now
$$\frac{dE}{dX} = 4 - \frac{2}{X^3(1+X)} - \frac{1}{X^2(1+X)^2},$$

$$X^3(1+X)^2\frac{dE}{dX} = 4X^3(1+X)^2 - 2(1+X) - X$$
$$= (2X^2-1)(2X^3+4X^2+3X+2) \geqslant 0 \quad \text{for} \quad X \geqslant 1/\sqrt{2}.$$

Hence E increases steadily.

LEMMA 19. *The line $U+V=2d$, where $\frac{1}{2} < d \leqslant t$, meets the curve $F(U, V) = 0$ in at most two points in the first quadrant if*

$$d^p > 1 - \tfrac{1}{2}p. \tag{8.29}$$

Put
$$U = d+\zeta, \quad V = d-\zeta, \tag{8.30}$$

so that
$$d+\zeta < 1, \quad \zeta < \tfrac{1}{2}, \tag{8.31}$$

Then

$$[(d+1+\zeta)^p - 1][(d+1-\zeta)^p - 1] - [1-(d+\zeta)^p][1-(d-\zeta)^p] = 0.$$

Write

$$g_1(\zeta) = [(d+1)^2 - \zeta^2]^p - (d+1+\zeta)^p - (d+1-\zeta)^p,$$
$$g_2(\zeta) = (d^2-\zeta^2)^p - (d+\zeta)^p - (d-\zeta)^p.$$

Then ζ satisfies the equation

$$g(\zeta) = g_1(\zeta) - g_2(\zeta) = 0.$$

Since $g(\zeta)$ is an even function of ζ, it suffices to show that $g(\zeta) = 0$ has at most one root in $0 < \zeta < 1-d$, or simply that $g'(\zeta) = 0$ has no root in $0 < \zeta < 1-d$. The equation $g'(\zeta) = 0$ can be written as

$$\phi(d+1) - \phi(d) = 0,$$

where

$$\phi(Z) = 2\zeta(Z^2-\zeta^2)^{p-1} + (Z+\zeta)^{p-1} - (Z-\zeta)^{p-1}.$$

Now $\phi(Z)$ is differentiable in Z for $d \leqslant Z \leqslant d+1$, since $\zeta < d$. Hence

$$\phi(d+1) - \phi(d) = \phi'(D),$$

for some D in $d < D < d+1$. Then

$$4\zeta D(D^2-\zeta^2)^{p-2} + (D+\zeta)^{p-2} - (D-\zeta)^{p-2} = 0,$$

or
$$h(\zeta) = 4\zeta D - (D+\zeta)^{2-p} + (D-\zeta)^{2-p} = 0.$$

Now $$h(\zeta) = h(0)+\zeta h'(0)+\tfrac{1}{2}\zeta^2 h''(\vartheta),$$

for some ϑ in $0<\vartheta<1$, and where, in evaluating $h(0)$, $h'(0)$, etc., D can be treated as a constant. Then

$$h''(\zeta) = (2-p)(1-p)[(D-\zeta)^{-p}-(D+\zeta)^{-p}] > 0,$$

$$h(0) = 0,$$

$$h'(0) = 4D-2(2-p)D^{1-p}.$$

Hence $h(\zeta)>0$ when $0<\zeta<D$ if $h'(0)>0$, i.e. if $D^p > 1-\frac{1}{2}p$, and this certainly holds if $d^p > 1-\frac{1}{2}p$.

On putting $2d = j+1/j-1$, we see that Theorem 5 is established whenever

$$j+1/j-1 > 2(1-\tfrac{1}{2}p)^{1/p}.$$

LEMMA 20. *If* $0<p\leqslant\frac{1}{3}$,

$$j+1/j-1 > 2(1-\tfrac{1}{2}p)^{1/p}.$$

It is easy to verify this inequality for a given value of p. Thus if $p=\frac{1}{3}$, it suffices to show that

$$j+1/j > 1+2(\tfrac{5}{6})^3 = \tfrac{233}{108} = 2\cdot157407\ldots,$$

or that $j<\cdot6742$, since

$$\cdot6742+1/\cdot6742 = \cdot6742+1\cdot48323\ldots = 2\cdot15743\ldots.$$

We have now to show that if

$$(1+j)^{\frac{1}{3}}+(1-j)^{\frac{1}{3}} = 1+j^{\frac{1}{3}},$$

then $j<\cdot6742$. In fact $j=\cdot672\ldots$. For if

$$g(z) = (1+z)^{\frac{1}{3}}+(1-z)^{\frac{1}{3}}-1-z^{\frac{1}{3}},$$

we have seen in (2.6) that $g(z)$ decreases steadily in $0\leqslant z\leqslant 1$. Also

$$g(\cdot672) = (1\cdot672)^{\frac{1}{3}}+(\cdot328)^{\frac{1}{3}}-1-(\cdot672)^{\frac{1}{3}},$$

$$> 1\cdot18689+\cdot68964-1-\cdot87591 = \cdot00062.$$

Next

$$g(\cdot673) = (1\cdot673)^{\frac{1}{3}}+(\cdot327)^{\frac{1}{3}}-1-(\cdot673)^{\frac{1}{3}},$$

$$< 1\cdot18714+\cdot68895-1-\cdot87633 = -\cdot00024.$$

It is not difficult to prove the result of Lemma 20 for $0 < p < \frac{1}{3}$ on using a table* of values of j, p calculated by Dr. Mahler and his result that j is a continuous increasing function of p in $0 < p < 1$. Thus I have proved it for successive intervals $0 < p \leqslant \cdot 1$, $\cdot 1 < p \leqslant \cdot 15$, $\cdot 15 < p \leqslant \cdot 2$, $\cdot 2 < p \leqslant \cdot 23$, etc., but it does not seem worth while giving details in view of the conjecture given in Lemma 14.

I prove finally that when $p = \frac{1}{2}, \frac{1}{3}$, the lattices (3.12), (3.13) are critical lattices, so that (8.4) is the best possible result.

For $p = \frac{1}{2}$, we use II of Lemma 14. It is easy to see that if

$$(1+j)^{\frac{1}{2}} + (1-j)^{\frac{1}{2}} = 1 + j^{\frac{1}{2}},$$

then $j > \frac{2}{3}$. It suffices to note that $5^{\frac{1}{2}} + 1 > 3^{\frac{1}{2}} + 2^{\frac{1}{2}}$. Also

$$f(5, 0) = \sqrt{5} > 2 > 2c^{\frac{1}{2}}.$$

For $p = \frac{1}{3}$, we use III of Lemma 14. Then

$$j = \cdot 672 \ldots > \tfrac{2}{3}, \quad f(8, 0) = 2 > 2c^{\frac{1}{3}}.$$

Also the condition (5.17) becomes

$$(3b-2a)^{\frac{1}{3}} + (2b+3a)^{\frac{1}{3}} \geqslant 2c^{\frac{1}{3}} = a^{\frac{1}{3}} + b^{\frac{1}{3}},$$

or

$$(3j-2)^{\frac{1}{3}} + (2j+3)^{\frac{1}{3}} \geqslant 1 + j^{\frac{1}{3}}. \tag{8.32}$$

This inequality is satisfied. Thus the left-hand side is greater than

$$(\cdot 016)^{\frac{1}{3}} + (4 \cdot 344)^{\frac{1}{3}} > \cdot 2519 + 1 \cdot 6316 = 1 \cdot 8835,$$

whilst the right-hand side is less than

$$1 + (\cdot 673)^{\frac{1}{3}} < 1 \cdot 8764.$$

We prove finally (5.18) which becomes now

$$(3-4j)^{\frac{1}{3}} + (3j+4)^{\frac{1}{3}} \geqslant 1 + j^{\frac{1}{3}}. \tag{8.33}$$

Since $\frac{2}{3} < j < \cdot 675$, this is obvious since the left-hand side of (8.33) is greater than

$$(\cdot 3)^{\frac{1}{3}} + 6^{\frac{1}{3}} > \cdot 6 + 1 \cdot 8 = 2 \cdot 4,$$

and the right-hand side is less than 2.

---

* Thus $p = 0$ $\cdot 1$, $\cdot 2$, $\cdot 3$, $\cdot 4$, $\cdot 5$, $\cdot 6$, $\cdot 7$, $\cdot 8$, $\cdot 9$, 1

$j = \cdot 618$ $\cdot 632$, $\cdot 648$, $\cdot 666$, $\cdot 687$, $\cdot 713$, $\cdot 744$, $\cdot 782$, $\cdot 832$, $\cdot 898$, 1

Mr. J. C. P. Miller also kindly computed a small table.

This completes the proof. It suggests that (8.4) is the best possible result for $1 > p \geqslant p_0$, where p_0 is a number nearly equal to $\frac{1}{3}$, and given by expressing the fact that the root of

$$(1+x)^{p_0}+(1-x)^{p_0} = 1+x^{p_0},$$

also satisfies $$(3x-2)^{p_0}+(2x+3)^{p_0} = 1+x^{p_0}.$$

Addendum. Since Watson proved the conjecture of Lemma 15 after this paper was written, I now reconsider the result of Theorem 5, and show at once that (8.4) is still the best possible result for at least $1 > p \geqslant \frac{4}{9}$.

The argument used when $p = \frac{1}{2}$ remains valid provided that $5^p \geqslant 2$, and $j \geqslant \frac{2}{3}$. The first inequality gives $p \geqslant \log 2/\log 5 = \cdot 43\ldots$, and then Mahler's table and result shows that $j > \frac{2}{3}$. Hence (8.4) is the best possible result for $p \geqslant \cdot 44$.

Suppose next that $\frac{2}{3} \leqslant j \leqslant \frac{5}{7}$. This is so from Mahler's table certainly if $\cdot 3 \leqslant p < \cdot 44$. The argument used when $p = \frac{1}{3}$ is still valid if

$$(3j-2)^p+(2j+3)^p \geqslant 1+j^p, \tag{8.34}$$

$$(3-4j)^p+(3j+4)^p \geqslant 1+j^p, \tag{8.35}$$

these inequalities corresponding to (8.32), (8.33). But since $j \leqslant \frac{5}{7}$, *i.e.* $3-4j \geqslant 3j-2$, (8.35) is satisfied when (8.34) is. Hence (8.4) is the best possible result when $\cdot 3 \leqslant p < \cdot 44$ provided* that p is such that (8.34) is satisfied, where j is of course defined by

$$(1+j)^p+(1-j)^p = 1+j^p.$$

9. *The case* $f(x, y) = xy(x+y)$.

THEOREM 6. *Every lattice of determinant* $\Delta > 0$ *has one of its points other than O in the region*

$$|x||y|(|x|+|y|) \leqslant 30(\Delta/13)^{\frac{3}{2}}. \tag{9.1}$$

This is the best possible result in the sense that, if the constant $30/13^{\frac{3}{2}}$ *is diminished, each of the critical lattices* (3.12), (3.13) *has an infinity of points lying on one of the coordinate axes of* x, y, *but no other points in the new region.*

The conditions of §2 are satisfied. It suffices to prove that the region

$$x \geqslant 0, \quad y \geqslant 0. \quad xy(x+y) \geqslant 1$$

---

* Watson (7) has shown since that this holds for $p > 0{\cdot}3295\ldots$, and so (8.4) is the best possible result for $0{\cdot}3296 < p < 1$.

is convex. This follows at once since the real points of inflexion of the cubic $xy(x+y)=1$ lie at infinity. I remark also that, if $n>0$, the family of curves

$$xy(x+y)^n=1$$

has no real, finite point of inflexion. It suffices to show that $d^2y/dx^2 \neq 0$. Now

$$\frac{1}{x}+\frac{1}{y}\frac{dy}{dx}+\frac{n}{x+y}\left(1+\frac{dy}{dx}\right)=0.$$

Differentiate again and put $d^2y/dx^2=0$. Then

$$\frac{1}{x^2}+\frac{1}{y^2}\left(\frac{dy}{dx}\right)^2+\frac{n}{(x+y)^2}\left(1+\frac{dy}{dx}\right)^2=0,$$

and this gives no real values of x, y.

The equations determining a, b, c become

$$ab(a+b)=2c^3, \tag{9.2}$$

$$(a^2-b^2)\,2a=2c^3, \tag{9.3}$$

$$a^2+b^2=2. \tag{9.4}$$

Hence $$2a^2-ab-3b^2=0,$$

i.e. $$2a-3b=0,$$

so that this is an instance of the special case when a/b is rational. Hence

$$a=3\sqrt{\tfrac{2}{13}},\quad b=2\sqrt{\tfrac{2}{13}},\quad ab=\tfrac{12}{13}, \tag{9.5}$$

$$c^3=\tfrac{30}{13}\sqrt{\tfrac{2}{13}}. \tag{9.6}$$

This gives $$c^6=1800/13^3,$$

$$c=\cdot967\ldots,\quad c^2=\cdot935\ldots,\quad c^3=\cdot905\ldots. \tag{9.7}$$

If we now proceed with the equations (3.6)–(3.11) and write

$$q=x(U+\tfrac{1}{2}),\quad r=y(V+\tfrac{1}{2}),$$

as in (8.13), we find

$$UV(U+1)(V+1)(U+V+1)+V^2+U^2-4UV-U-V=0,$$

and wish to prove for the relevant (x, y) that

$$U+V+1\geqslant 2/xy. \tag{9.8}$$

But while in § 8 more could be proved, namely that

$$U+V+1 \geqslant 2/ab,$$

this inequality, which here becomes

$$U+V \geqslant \tfrac{7}{6},$$

does not hold except at the points $(x, y) = (a, b)$, (b, a). Hence we shall have to prove the equivalent of the full inequality (9.8). We take it in the following form. Let $P(x, y)$ be any point on the boundary of $\mathcal{R}$ between (b, a) and (c, c). We have to prove that the line $-yX+xY = 2$ cuts off from $-XY(-X+Y) = 2c^3$ a secant in the second quadrant not exceeding $(x^2+y^2)^{\frac{1}{2}}$. Put

$$-X = (1+t)/y, \quad Y = (1-t)/x. \tag{9.9}$$

Then

$$\phi(t) \equiv \left(\frac{1-t^2}{xy}\right)\left(\frac{1+t}{y}+\frac{1-t}{x}\right)-2c^3 = 0. \tag{9.10}$$

Since $x < y$, and $xy < 1$ from $xy(x+y) = 2c^3$ and $c < 1$,

$$\phi(+\infty) = +\infty, \quad \phi(1) < 0, \quad \phi(0) > 0, \quad \phi(-1) < 0. \tag{9.11}$$

Hence there are roots t_1, t_2, t_3 of (9.10) such that

$$0 < t_1 < 1, \quad -1 < t_2 < 0, \quad 1 < t_3 < \infty.$$

We wish to prove that

$$\frac{(t_1-t_2)^2}{x^2}+\frac{(t_1-t_2)^2}{y^2} \leqslant x^2+y^2,$$

i.e.

$$t_1-t_2 \leqslant xy. \tag{9.12}$$

This condition is satisfied when $x = y = c$. For then, from (9.10), $t_1 = -t_2 = d$, say, where $1-d^2 = c^6$; and we have to show that $1 \leqslant c^6+\frac{1}{4}c^4$, or say $1 < c^6+\frac{1}{4}c^6$, *i.e.* $c^6 > \cdot 8$, *i.e.* $1800/13^3 > \cdot 8$, which is true.

The equation of the squared differences of the roots of the cubic

$$\phi(t) \equiv At^3+Bt^2+Ct+D = 0$$

is

$$F(z) \equiv z(A^2z+3AC-B^2)^2+27A^2D^2$$
$$-18ABCD+4AC^3+4DB^3-B^2C^2 = 0. \tag{9.13}$$

We prove now that $(t_1-t_2)^2$ is the smallest root of this equation. This is obvious since it is shown now that $t_3 > 5$. Clearly for the points (x, y) considered,

$$x/y \geqslant b/a = \tfrac{2}{3} \quad \text{or} \quad 3x \geqslant 2y.$$

Hence from (9.10), $\phi(5) < 0$ since

$$x+y+5(x-y) = 2(3x-2y) \geqslant 0,$$

and so the result follows from (9.11).

Since $F(0) < 0$, we have to express the condition $F(x^2y^2) \geqslant 0$. Write $\phi(t) = 0$ as

$$(1-t^2)\left(x+y+(x-y)\,t\right) = x^3y^3(x+y),$$

or
$$t^3\frac{x-y}{x+y}+t^2-t\frac{x-y}{x+y}+x^3y^3-1 = 0.$$

Hence it suffices if

$$F \equiv x^2y^2\left(\left(\frac{x-y}{x+y}\right)^2 x^2y^2-3\left(\frac{x-y}{x+y}\right)^2-1\right)^2+27\left(\frac{x-y}{x+y}\right)^2(x^3y^3-1)^2$$

$$+18\left(\frac{x-y}{x+y}\right)^2(x^3y^3-1)-4\left(\frac{x-y}{x+y}\right)^4+4(x^3y^3-1)-\left(\frac{x-y}{x+y}\right)^2 \geqslant 0,$$

when (x, y) lies between the points (b, a), (c, c) of the boundary, and where the equality sign holds when, and only when, $(x, y) = (b, a)$. Clearly

$$F = \left(\frac{x-y}{x+y}\right)^4\left((x^3y^3-3xy)^2-4\right)$$

$$+\left(\frac{x-y}{x+y}\right)^2(27x^6y^6-2x^4y^4-36x^3y^3+6x^2y^2+8)+4x^3y^3+x^2y^2-4.$$

$$= IZ^2+JZ+K,$$

say, where $Z = \left(\frac{x-y}{x+y}\right)^2$.

Now as $P(x, y)$ moves on the boundary of $\mathcal{R}$ from (b, a) to (c, c), clearly $x+y$ steadily decreases and so, since $xy(x+y) = 2c^3$ on the boundary, $u = xy$ steadily increases from ab to c^2. But then I, J, K are steadily increasing functions of xy and $I < 0$.

This is obvious for K.

For I,

$$I = (x^3y^3-3xy)^2-4 = (x^2y^2-1)^2(x^2y^2-4) = (u^2-1)^2(u^2-4),$$

and has a maximum at $u = 1$ and a minimum at $u^2 = 3$. Hence I steadily increases as u increases from 0 to 1, and so $I < 0$ for $\frac{12}{13} \leqslant u \leqslant c^2$.

Finally for J, it suffices to show that $\frac{1}{2u}\frac{dJ}{du} > 0$. But

$$\frac{1}{2u}\frac{dJ}{du} = 81u^4-4u^2-54u+6$$

$$> 81u^4-54u$$

$$> 81u(\tfrac{17}{21}\tfrac{28}{97}-\tfrac{54}{81}) > 0.$$

Now IZ^2 increases steadily since $Z^2 \geqslant 0$ and decreases steadily. Write

$$JZ = (J-\lambda)Z+\lambda Z,$$

where $$\lambda = 27c^{12}-2c^8-36c^6+6c^4+8$$

is the value of J at (c, c), and so $J-\lambda \leqslant 0$. As above, $(J-\lambda)Z$ steadily increases. The remaining terms in F are

$$K+\lambda Z = K+\lambda\left(1-\frac{4xy}{(x+y)^2}\right) = K+\lambda\left(1-\frac{x^3y^3}{c^6}\right),$$

and obviously steadily increase if $\lambda \leqslant 4c^6$, i.e.

$$E = 27c^{12}-2c^8-40c^6+6c^4+8 \leqslant 0.$$

Now $\frac{3}{4} < c^6 < \frac{18}{21}$, and so $27c^{12} < 24c^6$, $16c^6 > 12$. Then

$$E < -2c^8+6c^4-4 = (1-c^4)(2c^4-4) < 0.$$

Hence F steadily increases as P moves from (b, a) to (c, c), and since $F = 0$ at (b, a), $F \geqslant 0$.

To complete the proof it suffices to show that the only integer values of ξ, η for which

$$|3\xi-2\eta||2\xi+3\eta|\{|3\xi-2\eta|+|2\xi+3\eta|\} < 30, \qquad (9.14)$$

make the left-hand side zero. Suppose then that

$$|3\xi-2\eta||2\xi+3\eta| \neq 0.$$

Then we must clearly have

$$|3\xi-2\eta| \leqslant 4, \quad |2\xi+3\eta| \leqslant 4,$$

since (9.14) is not true if $|3\xi-2\eta| \geqslant 5$ or if $|2\xi+3\eta| \geqslant 5$. Hence we easily find

$$13|\xi| < 3(4)+2(4) \quad \text{or} \quad |\xi| \leqslant 1,$$

$$13|\eta| < 2(4)+3(4) \quad \text{or} \quad |\eta| \leqslant 1,$$

Hence $\xi = \pm 1$, $\eta = 0$, or $\xi = 0$, $\eta = \pm 1$, or $\xi = \pm 1$, $\eta = \pm 1$; but these do not satisfy (9.14).

10. *The case* $f(x, y) = xy(x^2+y^2)$.

THEOREM 7. *Every lattice of determinant* $\Delta > 0$ *has one of its points other than O in the region*

$$|x||y|(x^2+y^2) \leqslant 2\Delta^2/\sqrt{(17)}. \qquad (10.1)$$

This is the best possible result. There are now four critical lattices given by (3.12), (3.13), *and* (3.15) *and its image in the Y axis.*

By the linear transformation

$$\sqrt{2}\,x = x'+y', \quad \sqrt{2}\,y = x'-y',$$

of determinant -1, it is clear that the region (10.1) can be replaced by the region

$$|x^4-y^4| \leqslant 4\Delta^2/\sqrt{(17)}. \qquad (10.2)$$

The conditions of § 2 are all satisfied when $f(x, y) = xy(x^2+y^2)$. The convexity condition, i.e. the absence of finite points of inflexion, is more easily shown from (10.2). Thus if

$$y^4 = x^4+d,$$

$$y^3\frac{dy}{dx} = x^3.$$

Differentiate again and put $d^2y/dx^2 = 0$. Then

$$y^2(dy/dx)^2 = x^2 \quad \text{giving} \quad x^6 = x^2y^4.$$

By symmetry $x = 0$ does not give a point of inflexion, and $x = \pm y$ gives a point of inflexion at infinity.

The equations giving a, b, c are

$$ab(a^2+b^2) = 2c^4, \quad (a^2-b^2)(2a^2+2b^2) = 2c^4, \tag{10.3}$$

$$a^2+b^2 = 2.$$

Hence $$ab = c^4, \quad 2a^2-2b^2 = c^4,$$

and so $$4(c^4)^2+(\tfrac{1}{2}c^4)^2 = (a^2+b^2)^2 = 4,$$

$$c^8 = \tfrac{16}{17}. \tag{10.4}$$

Also $b = aj$ where $2j^2+j = 2$, or

$$4j = -1+\sqrt{(17)}. \tag{10.5}$$

Let now $P(x, y)$ be any point on the boundary of $\mathcal{R}$ lying between (b, a) and (a, b). We require the condition that the line $-Xy+Yx = 2$ meets the curve $-XY(X^2+Y^2) = 2c^4$ in points such that the intercepted segment does not exceed $(x^2+y^2)^{\frac{1}{2}}$. Put

$$-X = (1+t)/y, \quad Y = (1-t)/x.$$

Then

$$\left(\frac{1-t^2}{xy}\right)\left(\frac{(1+t)^2}{y^2}+\frac{(1-t)^2}{x^2}\right) = 2c^4. \tag{10.6}$$

Since it is obvious that there are only two real points of intersection, this equation will have only two real roots t_1, t_2 and clearly $|t_1| < 1$, $|t_2| < 1$, since $X < 0$, $Y > 0$. We have to show that

$$|t_1-t_2| \leqslant xy, \tag{10.7}$$

where $$xy(x^2+y^2) = 2c^4, \tag{10.8}$$

and $$b \leqslant x \leqslant a, \quad b \leqslant y \leqslant a. \tag{10.9}$$

In the particular case when $x = y = c$, $t_1 = -t_2 = t$, where

$$(1-t^2)(1+t^2) = c^8.$$

The condition $|2t| \leqslant c^2$ is now satisfied with the equality sign since this requires $16(1-c^8) = c^8$, which by (10.4) is satisfied.

If the condition (10.7) is satisfied with strict inequality by all the points of the boundary of $\mathcal{R}$ lying between (b, a) and (a, b) except the end and middle points, it is easy to see that the condition (3.16) for the

new critical cases is satisfied. For it becomes

$$|XY|(X^2+Y^2) \geqslant 2c^4,$$

where $$X,\ Y = c\xi + \left(-\tfrac{1}{2}c \mp \frac{1}{c}\right)\eta,$$

i.e. $$|(\xi-\tfrac{1}{2}\eta)^4 - \eta^4/c^8| \geqslant 1,$$

i.e. $$|(\xi-\tfrac{1}{2}\eta)^4 - 17\eta^4/16| \geqslant 1,$$

i.e. $$|(2\xi-\eta)^4 - 17\eta^4| \geqslant 16.$$

This is obviously satisfied when η is even. It holds also when η is odd, for then

$$\eta^4 \equiv 1 \pmod{16}, \quad (2\xi-\eta)^4 \equiv 1 \pmod{16},$$

and so the number inside the modulus sign is a multiple of 16.

The lattices (3.12), (3.13) are also critical lattices. Thus for (3.12) we require that, for all integers ξ, η, not both zero,

$$|(b\xi-a\eta)(a\xi+b\eta)|(a^2+b^2)(\xi^2+\eta^2) \geqslant 2c^4,$$

i.e. $$|ab(\xi^2-\eta^2)+(b^2-a^2)\,\xi\eta|(\xi^2+\eta^2) \geqslant c^4,$$

i.e., from (10.3),

$$|\xi^2-\eta^2-\tfrac{1}{2}\xi\eta|(\xi^2+\eta^2) \geqslant 1.$$

This is obviously true if $\xi^2+\eta^2 > 1$, and also if $\xi = \pm 1$, $\eta = 0$ or $\xi = 0$, $\eta = \pm 1$.

We return now to the equation (10.6), which we write as

$$(1-t^2)[(x^2+y^2)(1+t^2)+(x^2-y^2)\,2t] = 2c^4x^3y^3,$$

i.e.

$$(1-t^2)\left(1+t^2+2\frac{x^2-y^2}{x^2+y^2}\,t\right) = 2c^4x^3y^3/(x^2+y^2) = 4c^8x^2y^2/(x^2+y^2)^2.$$

Write $$z = 2xy/(x^2+y^2), \tag{10.10}$$

so that from (10.9) and (10.3)

$$1 \geqslant z \geqslant 4/\sqrt{(17)} = c^4. \tag{10.11}$$

Also $$(1-t^2)\{1+t^2+2t(1-z^2)^{\frac{1}{2}}\} = c^8z^2. \tag{10.12}$$

Since from (10.10), (10.8), $x^2y^2 = c^4z$, we have to show that

$$|t_1-t_2| \leqslant 2\sqrt{z}/\sqrt[4]{(17)} = c^2\sqrt{z}, \tag{10.13}$$

or

$$|t_1-t_2|^2 \leqslant c^4z.$$

Let $F(T) = 0$ be the equation of squared differences of the roots of the quartic in t, (10.12). It will have one positive, one negative root and two pairs of conjugate complex roots. If, as we may suppose, the coefficient of T^6 is positive, the constant term will be negative. Hence (10.13) becomes $F(c^4z) \geqslant 0$ for all the values of z with $1 \geqslant z \geqslant c^4$.

Write the equation in t as

$$At^4+4Bt^3+6Ct^2+4Dt+E = 0.$$

Put

$$H = AC-B^2, \quad I = AE-4BD+3C^2, \quad J = ACE+2BCD-AD^2-EB^2-C^3.$$

Then the equation $F(T) = 0$ is given by M. Roberts (8) as

$$A^6T^6+48A^2HT^5+8A^2(96H^2+A^2I)T^4$$
$$+32(128H^3+16A^2HI-13A^3J)T^3+16(384H^2I-7A^2I^2-288AHJ)T^2$$
$$+1152(2HI-3AJ)IT+256(I^3-27J^2) = 0.$$

Since our equation is

$$t^4+2(1-z^2)^{\frac{1}{2}}t^3-2(1-z^2)^{\frac{1}{2}}t+c^8z^2-1 = 0,$$

$$A = 1, \quad B = -D = \tfrac{1}{2}(1-z^2)^{\frac{1}{2}}, \quad C = 0, \quad E = c^8z^2-1,$$

$$H = -(1-z^2)/4, \quad I = c^8z^2-1+1-z^2 = (c^8-1)z^2,$$

$$J = -(1-z^2)/4-(1-z^2)(c^8z^2-1)/4 = -c^8z^2(1-z^2)/4.$$

The condition $F(c^4z) \geqslant 0$ becomes

$$c^{24}z^6-12(1-z^2)(c^{20}z^5)+8c^{16}z^4[6(1-z^2)^2+(c^8-1)z^2]$$
$$+32c^{12}z^3[-2(1-z^2)^3-4z^2(1-z^2)(c^8-1)+13c^8z^2(1-z^2)/4]$$
$$+16c^8z^2[24(1-z^2)^2(c^8-1)z^2-7(c^8-1)^2z^4-18(1-z^2)^2c^8z^2]$$
$$+1152c^4z[-\tfrac{1}{2}(1-z^2)(c^8-1)z^2+3(1-z^2)c^8z^2/4][c^8-1]z^2$$
$$+256[(c^8-1)^3z^6-27c^{16}z^4(1-z^2)^2/16] \geqslant 0.$$

All the terms in this are divisible by $(1-z^2)$ except apparently the term

$$z^6[c^{24}+8c^{16}(c^8-1)-112c^8(c^8-1)^2+256(c^8-1)^3].$$

Since the inequality becomes an equality when $z=1$, this term must be zero. It is easily verified that the term is

$$z^6[c^8+16(c^8-1)][c^8-4(c^8-1)]^2=0,$$

since $17c^8=16$. Divide out by $z^3(1-z^2)$; then we have to prove that

$$-12c^{20}z^2+8c^{16}z(6-6z^2)+32c^{12}[-2(1-z^2)^2-4z^2(c^8-1)+13c^8z^2/4]$$
$$+16c^8[24z(1-z^2)(c^8-1)-18c^8z(1-z^2)]$$
$$+1152c^4[\tfrac{1}{2}(1-c^8)+3c^8/4][c^8-1]z^2-432c^{16}z[1-z^2]\geqslant 0.$$

Arrange the coefficients of the powers of z as shown by the scheme

z^4	z^3	z^2	z	1
$-64c^{12}$	$-48c^{16}$	$-12c^{20}$	$48c^{16}$	$-64c^{12}$
	$-384c^8(c^8-1)$	$128c^{12}$	$384(c^8-1)c^8$	
	$288c^{16}$	$-128c^{12}(c^8-1)$	$-288c^{16}$	
	$432c^{16}$	$104c^{20}$	$-432c^{16}$	
		$288c^4(c^8-1)(c^8+2)$		

We know that the z polynomial has a factor $z-c^4$, and it is clear on writing $-1/z$ for z in the polynomial that c^4z+1 is a factor. Write now $z-1/z=Z$. Then on dividing by z^2, we require that

$$-64c^{12}(Z^2+2)+(288c^{16}+384c^8)Z+252c^{20}+544c^{12}-576c^4\geqslant 0.$$

Since one factor is $\dfrac{1}{z}(z-c^4)(c^4z+1)=c^4Z+1-c^8$, the other factor is $-64c^8Z+\lambda$, where, by equating terms in Z,

$$\lambda c^4-64c^8+64c^{16}=288c^{16}+384c^8,$$

or

$$\lambda=224c^{12}+448c^4.$$

It is easy to verify the factorisation, which requires that

$$\lambda(1-c^8)=252c^{20}+416c^{12}-576c^4,$$

i.e. $$(1-c^8)(112+56c^8) = 63c^{16}+104c^8-144,$$

i.e. $$119c^{16}+160c^8-256 = 0,$$

i.e. $$(17c^8-16)(7c^8+16) = 0,$$

which holds by (10.4).

The second factor

$$\lambda-64c^8 Z \quad \text{i.e.} \quad 224c^{12}+448c^4-64c^8 Z$$

is clearly positive since, for $c^4 \leqslant z \leqslant 1$, $c^4-1/c^4 \leqslant Z \leqslant 0$.

The first factor is also positive.

This proves (10.13) and completes the proof of Theorem 7.

11. *Postscript.*

It has been suggested to me that it might be of interest to apply my method to the well-known case mentioned in the introduction when

$$f(x, y) = xy. \tag{11.1}$$

I prove

THEOREM 8. *Every lattice of determinant $\Delta > 0$ has one of its points other than O in the region*

$$|xy| \leqslant \Delta/\sqrt{5}. \tag{11.2}$$

This is the best possible result.

There are two infinite sets of critical lattices of which one is given by

$$X = \lambda(\Delta/\sqrt{5})^{\frac{1}{2}}\left(\xi-\tfrac{1}{2}(1+\sqrt{5})\eta\right), \quad Y = \frac{1}{\lambda}(\Delta/\sqrt{5})^{\frac{1}{2}}\left(\xi-\tfrac{1}{2}(1-\sqrt{5})\eta\right), \tag{11.3}$$

where λ is any real number.

Proceeding as in § 3, we have

$$ab = a^2-b^2 = c^2, \tag{11.4}$$

$$a^2+b^2 = 2, \tag{11.5}$$

whence $$(2+c^2)(2-c^2) = 4a^2b^2 = 4c^4,$$

i.e. $$c^4 = \tfrac{4}{5}. \tag{11.6}$$

Then (11.2) agrees with $xy \leqslant c^2$ when $\Delta = 2$.

We have now to express the condition that the line $-yX+xY=2$ cuts off from $-XY=c^2$ a secant in the second quadrant not exceeding $(x^2+y^2)^{\frac{1}{2}}$. Put

$$-X=(1+t)/y, \quad Y=(1-t)/x,$$

so that

$$1-t^2=c^2xy=c^4,$$

i.e.

$$t_1=1/\sqrt{5}, \quad t_2=-1/\sqrt{5}. \tag{11.7}$$

The condition $t_1-t_2\leqslant xy$ as in (9.12), say, becomes

$$2t_1\leqslant c^2, \tag{11.8}$$

which is true with the equality sign. Hence any point $P(x, y)$ on the arc of $xy=c^2$ between (b, a) and (a, b) may, and in fact does, lead to two critical lattices. Thus we have from (3.10) the possible critical lattice

$$X=x\xi+(-q-\tfrac{1}{2}x)\eta, \quad Y=y\xi+(r-\tfrac{1}{2}y)\eta. \tag{11.9}$$

Since $2q=(1+t_1)/y+(1-t_1)/y$, $2r=(1-t_1)/x+(1+t_1)/x$, we have

$$q=1/y, \quad r=1/x.$$

Then (11.9), which takes the form

$$yX=c^2(\xi-\tfrac{1}{2}\eta)-\eta, \quad xY=c^2(\xi-\tfrac{1}{2}\eta)+\eta,$$

is a critical lattice if

$$|c^2(\xi-\tfrac{1}{2}\eta)-\eta||c^2(\xi-\tfrac{1}{2}\eta)+\eta|\geqslant c^4$$

for all integers ξ, η except 0, 0; *i.e.*

$$|c^4(\xi-\tfrac{1}{2}\eta)^2-\eta^2|\geqslant c^4,$$

i.e.

$$|(2\xi-\eta)^2-5\eta^2|\geqslant 4,$$

i.e.

$$|\xi^2-\xi\eta-\eta^2|\geqslant 1,$$

which is true.

In (11.9), (x, y) lies on the arc of $xy=c^2$ between (b, a) and (a, b), but writing (11.9) in the form (11.3), the result still holds without any restriction on λ.

References.

1. Mordell, L. J. "On numbers representable by binary cubic forms", *Proc. London Math. Soc.* (2), 48 (1944), 198–228.
2. Korkine, A., and Zolotareff, G. "Sur les formes quadratiques", *Math. Annalen*, 6 (1873), 366–389, 369–370.
3. Markoff, A. "Sur les formes quadratiques binaires indéfinies", *Math. Annalen*, 15 (1879), 381–406; 17 (1880), 379–399.
4. Davenport, H. "On the product of three homogeneous linear forms (III)", *Proc. London Math. Soc.* (2), 45 (1939), 98–125.
5. Minkowski, H. *Diophantische Approximationen* (Berlin, 1907), 47–58.
6. Watson, G. N. "Proof of a conjecture stated by Mordell", *Journal London Math. Soc.*, 16 (1941), 157–166.
7. ———— "A problem in elementary analysis proposed by Mordell", *Proc. London Math. Soc.* (2), 48 (1944), 391–400.
8. Roberts, M., quoted in Burnside and Panton, *Theory of equations*, 1 (1899), 144.

University,
Manchester.

A PROBLEM IN ELEMENTARY ANALYSIS PROPOSED BY MORDELL

By G. N. WATSON.

[Received 27 April, 1942.—Read 21 May, 1942.]

1. A crucial result arising in a lattice-point problem in the preceding paper by Prof. Mordell† may be enunciated in the following form‡:

Let

$$f(x, y) \equiv (1+y)^x + (1-y)^x - y^x - 1 \quad (0 \leqslant x \leqslant 1;\ 0 \leqslant y \leqslant 1),$$

$$F(x, y) \equiv (2y+3)^x + (3y-2)^x - y^x - 1 \quad (0 \leqslant x \leqslant 1;\ \tfrac{2}{3} \leqslant y \leqslant 1).$$

Then the curves $f(x, y) = 0$, $F(x, y) = 0$ *have one and only one point of intersection,* (ξ, η) *say, inside the rectangle whose sides are* $x = 0$, $x = 1$, $y = \frac{2}{3}$, $y = 1$.

Moreover, if (ξ', η') *is any point on the curve* $f(x, y) = 0$ *and inside the rectangle, with* $\xi' > \xi$, *then* $F(\xi', \eta')$ *is positive.*

The present paper contains a proof of this conjecture. I also give an investigation of the shape of the curve $F^*(x, y) = 0$, where

$$F^*(x, y) \equiv (2y+3)^x - (2-3y)^x - y^x - 1 \quad (0 \leqslant x \leqslant 1;\ 0 \leqslant y \leqslant \tfrac{2}{3}),$$

similar to my work on the curves $f(x, y) = 0$ and $F(x, y) = 0$. The connexion between the second and third of the curves will be explained in § 5.

I must explain here that, before Prof. Mordell called my attention to the need for proving the conjecture, Dr. K. Mahler had studied the topography of the first of the curves (but not the second or third), and he had

† L. J. Mordell, *Proc. London Math. Soc.* (2), 48 (1944), 339–390 (378).

‡ It is convenient to use the terminology of two-dimensional analytical geometry.

obtained the results about it which are given in § 2 below; I therefore began by studying the topography of the second curve, and then dealt with the first curve in the same way. When I subsequently compared Dr. Mahler's work with my own, it seemed to me that my procedure was slightly briefer and more direct (though my opinion may be merely paternal prejudice), and accordingly I give only my own work in § 2 without introducing any of Dr. Mahler's analysis.

2. Consider the function $f(x, y)$; it is a continuous function of both variables throughout the "unit square" $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, except at the vertices (0, 0) and (0, 1). On the side $x = 0$ (with the exception of these vertices) we have $f(0, y) = 0$. On the opposite side $x = 1$ we have $f(1, y) = 1-y$, and this is positive except at (1, 1), where it is zero. On the side $y = 0$ (except at the origin) we have $f(x, 0) = 1$. On the opposite side $y = 1$, the point (0, 1) excepted, we have $f(x, 1) = 2^x-2$, and this is negative except at (1, 1), where it is zero.

Next, for any fixed x $(0 < x < 1)$ and $0 < y < 1$, we have

$$\frac{\partial f(x, y)}{\partial y} = x\left\{\frac{1}{(1+y)^{1-x}} - \frac{1}{(1-y)^{1-x}}\right\} - xy^{x-1}$$

$$< -xy^{x-1} < 0.$$

Hence, as (x, y) moves along any fixed ordinate, $x = h$ say, from $(h, 0)$ up to $(h, 1)$, $f(x, y)$ steadily decreases from the positive value 1 to the negative value 2^h-2, and so it passes through the value zero once and only once. That is to say, the curve $f(x, y) = 0$ meets each ordinate $x = h$ $(0 < h < 1)$ once and only once.

It is easy to see that this point of intersection is above the line $y = \frac{1}{2}$, since $f(x, \frac{1}{2}) = (\frac{3}{2})^x-1 > 0$; but we can get more precise information than this in the following manner: let y_0 be the positive root of the equation

$$1-y^2 = y,$$

so that $$y_0 = \tfrac{1}{2}(-1+\sqrt{5}) = 0{\cdot}61803\ldots.$$

Then, for $0 < x \leqslant 1$, we have

$$f(x, y_0) = (1+y_0)^x+(1-y_0)^x-(1-y_0^2)^x-1$$

$$= \{(1+y_0)^x-1\}\{1-(1-y_0)^x\} > 0;$$

and therefore the whole of the curve $f(x, y) = 0$, apart from the left-hand side of the unit square, must lie above the line $y = y_0$.

Also, for small positive values of x with $0<y<1$, we have

$$f(x, y) = x \log \frac{1-y^2}{y} + O(x^2),$$

and so, if (x, y) moves along the curve $f(x, y) = 0$, we see that $y \to y_0$ as $x \to +0$.

We next consider how $f(x, y)$ behaves when x is varied, y remaining fixed. We have, for any fixed y $(y_0 < y \leqslant 1)$ and $0 < x \leqslant 1$,

$$\frac{\partial f(x, y)}{\partial x} = (1+y)^x \log (1+y) + (1-y)^x \log (1-y) - y^x \log y$$

$$= y^x \phi(x, y),$$

say, where

$$\phi(x, y) \equiv \left(\frac{1+y}{y}\right)^x \log (1+y) + \left(\frac{1-y}{y}\right)^x \log (1-y) - \log y,$$

$$\frac{\partial \phi(x, y)}{\partial x} = \left(\frac{1+y}{y}\right)^x \log\left(\frac{1+y}{y}\right) \log (1+y) + \left(\frac{1-y}{y}\right)^x \log\left(\frac{1-y}{y}\right) \log (1-y).$$

The first term on the right in the last formula is evidently positive, and the second term is also positive when $y_0 < y < 1$, since both of the logarithmic factors are then negative. Hence, along the line $y = k$ $(y_0 < k < 1)$, $\phi(x, k)$ is an increasing function of x. It is negative when $x = 0$, since

$$\phi(0, k) = \log \frac{1-k^2}{k} < \log \frac{1-y_0^2}{y_0} = 0;$$

but it cannot be negative all the way along the line; for, if it were, $f(x, k)$ would be a decreasing function of x all along the line, whereas

$$f(1, k) = 1-k > 0 = f(0, k).$$

Hence, as x increases from 0 to 1, $\phi(x, k)$ is first negative and subsequently positive; that is to say, $f(x, k)$ decreases from zero to a (negative) minimum and then increases up to the positive value $1-k$. It therefore vanishes for one and only one value of x between 0 and 1; and, for this value of x, it is increasing, so that $\partial f(x, k)/\partial x$ is positive.

Now the slope of the curve $f(x, y) = 0$ is given by the formula

$$\frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} \frac{dy}{dx} = 0;$$

and so, since at all points on the curve inside the unit square we have

$$\frac{\partial f(x, y)}{\partial x} > 0, \quad \frac{\partial f(x, y)}{\partial y} < 0,$$

the curve has a continuous positive slope. The part of the curve inside the unit square consequently consists of an arc with positive slope, the arc steadily rising and moving to the right from $(0, y_0)$ to $(1, 1)$. The left-hand side of the unit square (with end-points omitted) also satisfies the equation of the curve, but no other points, except $(1, 1)$, on the boundary of the square lie on the curve.

This completes the investigation of the curve $f(x, y) = 0$. Its shape is shown in Fig. 1 below.

3. We next consider the function $F(x, y)$; with the exception of the point $(0, \frac{2}{3})$, this function is a continuous function of both variables throughout the rectangle R defined by the inequalities $0 \leqslant x \leqslant 1$, $\frac{2}{3} \leqslant y \leqslant 1$. On the side $x = 0$ (with the exception of the above-mentioned point) we have $F(0, y) = 0$. On the opposite side $x = 1$ we have $F(1, y) = 4y$, and this is positive. On the side $y = \frac{2}{3}$ (with the exception of the above-mentioned point) we have

$$F(x, \tfrac{2}{3}) = (\tfrac{1}{3}\tfrac{3}{3})^x - (\tfrac{2}{3})^x - 1;$$

we denote the function on the right by $\Psi(x)$, and we shall investigate it in a moment. On the opposite side $y = 1$ we have $F(x, 1) = 5^x - 1$, and this is positive except at the end-point $(0, 1)$.

We now observe that, for $0 \leqslant x \leqslant 1$, $\Psi(x)$ has the positive differential coefficient

$$(\tfrac{1}{3}\tfrac{3}{3})^x \log \tfrac{1}{3}\tfrac{3}{3} + (\tfrac{2}{3})^x \log \tfrac{3}{2},$$

while $\Psi(0) = -1$, $\Psi(1) = \frac{8}{3}$. Hence $\Psi(x)$ vanishes for one and only one value of x between 0 and 1. We call this value x_2, and then we have $F(x, \frac{2}{3}) < 0$ or $F(x, \frac{2}{3}) > 0$ according as $x < x_2$ or $x > x_2$. The locus $F(x, y) = 0$ consequently consists in part of the left-hand side of the rectangle R, but it does not meet the other three sides of the rectangle except at the point $(x_2, \frac{2}{3})$. I have computed the value of x_2 to ten places of decimals, my result being

$$x_2 = 0{\cdot}41743\,24316.$$

Next, for any fixed x $(0<x<1)$ and $\frac{2}{3}<y<1$, we have

$$\frac{\partial F(x,y)}{\partial y} = x\left\{\frac{2}{(2y+3)^{1-x}}+\frac{3}{(3y-2)^{1-x}}-\frac{1}{y^{1-x}}\right\}$$

$$> \frac{2x}{(2y+3)^{1-x}}+x\left\{\frac{1}{(y-\frac{2}{3})^{1-x}}-\frac{1}{y^{1-x}}\right\},$$

and, for $0<x\leqslant 1$, this is positive for all values of y exceeding $\frac{2}{3}$, whether less than 1 or not. Hence, as (x, y) moves along any fixed ordinate, $x=h$ say, from $(h, \frac{2}{3})$ up to (h, ∞), $F(x, y)$ increases, starting from the negative value $\Psi(h)$ when $0<h<x_2$, but starting from the positive value $\Psi(h)$ when $x_2<h\leqslant 1$; in both cases it passes through the positive value 5^h-1 when $h=1$ and assumes larger positive values as y continues to increase up to ∞. That is to say, the curve $F(x, y)=0$ meets each fixed ordinate $x=h$ either once and only once or else not at all, according as $0<h<x_2$ or $x_2<h\leqslant 1$; and, if there is an intersection, it lies below the line $y=1$.

We can get a bound lower than 1 to the ordinate of the point of intersection in the following manner: let y_1 be the positive root of the equation

$$(2y+3)(3y-2)=y,$$

so that $$y_1=\tfrac{1}{3}(-1+\sqrt{10})=0{\cdot}72075\,9\ldots.$$

Then, for $0<x\leqslant 1$, we have

$$F(x, y_1) = (2y_1+3)^x+(3y_1-2)^x-(2y_1+3)^x(3y_1-2)^x-1$$

$$= \{(2y_1+3)^x-1\}\{1-(3y_1-2)^x\}>0,$$

and therefore the whole of the curve $F(x, y)=0$, apart from the y-axis, must lie below the line $y=y_1$.

Also, for small positive x with $\frac{2}{3}<y<1$, we have

$$F(x, y) = x\log\frac{(2y+3)(3y-2)}{y}+O(x^2),$$

and so, if (x, y) moves along the curve $F(x, y)=0$, we see that $y\to y_1$ as $x\to+0$.

We next consider how $F(x, y)$ behaves when x is varied, y remaining fixed. We have, for any fixed y $(\frac{2}{3}<y<y_1)$ and $0<x\leqslant 1$,

$$\frac{\partial F(x, y)}{\partial x} = (2y+3)^x\log(2y+3)+(3y-2)^x\log(3y-2)-y^x\log y$$

$$= y^x\Phi(x, y)$$

say, where

$$\Phi(x, y) \equiv \left(\frac{2y+3}{y}\right)^x \log(2y+3) + \left(\frac{3y-2}{y}\right)^x \log(3y-2) - \log y,$$

$$\frac{\partial \Phi(x, y)}{\partial x} = \left(\frac{2y+3}{y}\right)^x \log\left(\frac{2y+3}{y}\right) \log(2y+3)$$

$$+ \left(\frac{3y-2}{y}\right)^x \log\left(\frac{3y-2}{y}\right) \log(3y-2).$$

The first term on the right in the last formula is evidently positive, and the second term is also positive because $3y-2 < y$ and $3y-2 < 1$, so that both of the logarithmic factors are negative. Hence, along the line $y = k$ $(\frac{2}{3} < k < y_1)$, $\Phi(x, k)$ is an increasing function of x. It is negative when $x = 0$, since

$$\Phi(0, k) = \log \frac{(2k+3)(3k-2)}{k} < \log \frac{(2y_1+3)(3y_1-2)}{y_1} = 0,$$

but it cannot be negative all the way along the line; for, if it were, $F(x, k)$ would be a decreasing function of x all along the line, whereas

$$F(1, k) = 4k > 0 = F(0, k).$$

Hence, as x increases from 0 to 1, $\Phi(x, k)$ is first negative and subsequently positive; that is to say, $F(x, k)$ decreases from zero to a (negative) minimum and then increases up to the positive value $4k$. It therefore vanishes for one and only one value of x between 0 and 1; and, for this value of x, it is increasing, so that $\partial F(x, k)/\partial x$ is positive.

Now the slope of the curve $F(x, y) = 0$ is given by the formula

$$\frac{\partial F(x, y)}{\partial x} + \frac{\partial F(x, y)}{\partial y} \frac{dy}{dx} = 0$$

and so, since at all points on the curve inside the rectangle R both of the partial differential coefficients $\partial F(x, y)/\partial x$ and $\partial F(x, y)/\partial y$ are positive, the curve must have a continuous negative slope. The part of the curve inside the rectangle R consequently consists of an arc with negative slope, the arc steadily falling and moving to right from $(0, y)$ to $(x_2, \frac{2}{3})$. The y-axis above the point $(0, \frac{2}{3})$ also satisfies the equation of the curve, but no other points on the boundary of the rectangle R lie on the curve.

This completes the investigation of the curve $F(x, y) = 0$. Its shape is shown by a continuous line in Fig. 2.

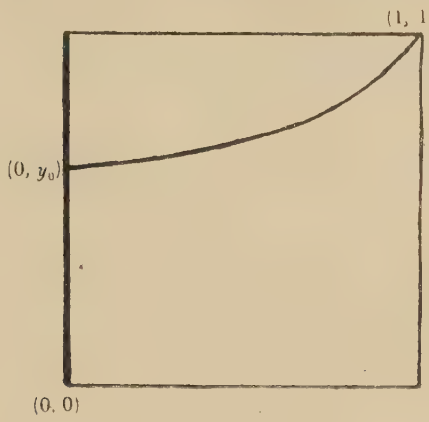


Fig. 1. $f(x, y) = 0$.

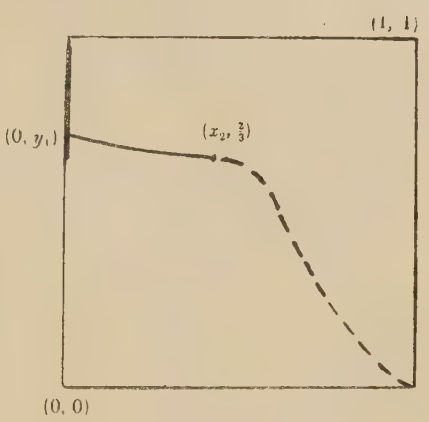


Fig. 2. $F(x, y) = 0$; $F^*(x, y) = 0$.

4. The proof of the theorem enunciated by Mordell is now quite brief. We first compute the value of $f(x_2, \frac{2}{3})$, the result to six significant figures being

$$f(x_2, \tfrac{2}{3}) = +0{\cdot}02555\,18.$$

Since $f(x_2, y)$ is a decreasing function of y, it follows from this result that the curve $f(x, y) = 0$ meets the ordinate $x = x_2$ at a point above $(x_2, \frac{2}{3})$. Hence, as x increases from 0 to x_2, the ordinate of the curve $F(x, y) = 0$ steadily decreases from y_1 to $\frac{2}{3}$, while the ordinate of the curve $f(x, y) = 0$. steadily increases from y_0 (a value less than y_1) to a value exceeding $\frac{2}{3}$. The curves therefore intersect at one and only one point, (ξ, η) say. This is the first part of Mordell's conjecture.

I have computed the values of ξ and η to ten places of decimals, my results being

$$\xi = 0{\cdot}32952\,85263, \quad \eta = 0{\cdot}67191\,58318,$$

Finally let (ξ', η') be any point on the curve $f(x, y) = 0$ such that $\xi' > \xi$. Since the curve has a positive slope, we consequently have $\eta' > \eta$. Now $F(\xi', y)$ is an increasing function of y throughout the rectangle R, while $F(x, \eta)$ is an increasing function of x when $\xi \leqslant x \leqslant 1$. We therefore have immediately

$$F(\xi', \eta') > F(\xi', \eta) > F(\xi, \eta) = 0,$$

and this completes the proof of Mordell's conjecture.

5. Finally we study the curve $F^*(x, y) = 0$ for $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant \frac{2}{3}$. The relevance of this is that, if (x, y) is a point on this curve and if x is a rational fraction which, when expressed in its lowest terms, has an odd denominator, then (x, y) is a point on the curve $F(x, y) = 0$. That is to say, the curve $F(x, y) = 0$ has a set of conjugate points lying along the curve $F^*(x, y) = 0$.

The curve $F^*(x, y) = 0$ is slightly more troublesome to investigate than the two previous curves, since it seems necessary to make use of higher differential coefficients.

Consider now the function $F^*(x, y)$; with the exception of the points $(0, 0)$ and $(0, \frac{2}{3})$, this function is a continuous function of both variables throughout the rectangle R^* defined by the inequalities $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant \frac{2}{3}$. On the side $x = 0$ (with the exception of the end points) we have $F^*(0, y) = -2$. On the opposite side $x = 1$ we have $F^*(1, y) = 4y$, and this is positive. On the side $y = 0$ (with the exception of the origin) we have

$$F^*(x, 0) = 3^x - 2^x - 1.$$

The function on the right is an increasing function of x in the interva $(0, 1)$, and it vanishes when $x = 1$. Hence $F^*(x, y)$ is negative when $y = 0$, $0 < x < 1$. On the opposite side, $y = \frac{2}{3}$, of the rectangle R^* we have $F^*(x, \frac{2}{3}) = \Psi(x)$, and so $F^*(x, \frac{2}{3})$ is negative or positive according as $0 < x < x_2$ or $x_2 < x \leqslant 1$.

Next, for any fixed x $(0 < x < 1)$ and $0 \leqslant y \leqslant \frac{2}{3}$, we have

$$\frac{\partial F^*(x, y)}{\partial y} = x\left\{\frac{2}{(2y+3)^{1-x}} + \frac{3}{(2-3y)^{1-x}} - \frac{1}{y^{1-x}}\right\},$$

$$\frac{\partial^2 F^*(x, y)}{\partial y^2} = x(1-x)\left\{-\frac{4}{(2y+3)^{2-x}} + \frac{9}{(2-3y)^{2-x}} + \frac{1}{y^{2-x}}\right\}$$

$$> x(1-x)\left\{\frac{4}{(2-3y)^{2-x}} - \frac{4}{(2y+3)^{2-x}}\right\} > 0.$$

Hence $\partial F^*(x, y)/\partial y$ is an increasing function of y, and its values for $y = 0$ and $y = \frac{2}{3}$ are $-\infty$ and $+\infty$ respectively. Consequently, as y increases from 0 to $\frac{2}{3}$, x remaining fixed, $F^*(x, y)$ first decreases from the negative value $F^*(x, 0)$ to a (negative) minimum and then increases up to the value $\Psi(x)$, which is negative or positive according as $0 < x < x_2$ or $x_2 < x < 1$. That is to say, the curve $F^*(x, y) = 0$ meets any fixed ordinate $x = h$ $(0 < h < 1)$ either not at all or else once and only once, according as $0 < h < x_2$ or $x_2 < h < 1$; and, if there is an intersection, $\partial F^*(x, y)/\partial y$ is positive there.

We next consider how $F^*(x, y)$ behaves when x is varied, y remaining fixed. We have, for any fixed y $(0 < y < \frac{2}{3})$ and $0 < x < 1$,

$$\frac{\partial F^*(x, y)}{\partial x} = (2y+3)^x \log (2y+3) - (2-3y)^x \log (2-3y) - y^x \log y$$
$$= y^x \Phi^*(x, y)$$

say, where

$$\Phi^*(x, y) \equiv \left(\frac{2y+3}{y}\right)^x \log (2y+3) - \left(\frac{2-3y}{y}\right)^x \log (2-3y) - \log y,$$

$$\frac{\partial \Phi^*(x, y)}{\partial x} = \left(\frac{2y+3}{y}\right)^x \log \left(\frac{2y+3}{y}\right) \log (2y+3)$$
$$- \left(\frac{2-3y}{y}\right)^x \log \left(\frac{2-3y}{y}\right) \log (2-3y)$$
$$= \left(\frac{2-3y}{y}\right)^x \chi(x, y)$$

say, where

$$\chi(x, y) \equiv \left(\frac{2y+3}{2-3y}\right)^x \log \left(\frac{2y+3}{y}\right) \log (2y+3) - \log \left(\frac{2-3y}{y}\right) \log (2-3y),$$

so that, for y fixed $(0 < y < \frac{2}{3})$, $\chi(x, y)$ is obviously an increasing function of x.

We next examine the value of $\chi(0, y)$. For $0 < y \leqslant \frac{1}{3}$ we have

$$\log \left(\frac{2y+3}{y}\right) > \log \left(\frac{2-3y}{y}\right) > 0, \quad \log (2y+3) > \log (2-3y) \geqslant 0,$$

so that $\chi(0, y)$ is then positive. For $\frac{1}{3} < y < \frac{1}{2}$ we have

$$\chi(0, y) = \log \left(\frac{2y+3}{y}\right) \log (2y+3) + \log \left(\frac{2-3y}{y}\right) \log \frac{1}{2-3y},$$

all four of the logarithms being positive, so that $\chi(0, y)$ is still positive. For $\frac{1}{2} \leqslant y < \frac{2}{3}$ we have

$$\chi(0, y) = \log \left(\frac{2y+3}{y}\right) \log (2y+3) - \log \left(\frac{y}{2-3y}\right) \log \left(\frac{1}{2-3y}\right).$$

In the expression on the right we have

$$\log \left(\frac{2y+3}{y}\right) > \log \left(\frac{1}{2-3y}\right) > 0, \quad \log (2y+3) > \log \left(\frac{y}{2-3y}\right) \geqslant 0$$

for $\frac{1}{2} \leqslant y < \frac{1}{2}(-1+\sqrt{5}) = y_0$; on the other hand we have

$$\log\left(\frac{1}{2-3y}\right) > \log\left(\frac{2y+3}{y}\right) > 0, \quad \log\left(\frac{y}{2-3y}\right) > \log(2y+3) > 0$$

for $y_0 < y < \frac{2}{3}$. Thus $\chi(0, y)$ is positive for $\frac{1}{2} \leqslant y < y_0$, but it is negative for $y_0 < y < \frac{2}{3}$.

Hence, collecting our results, we have

$$\chi(0, y) > 0 \quad (0 < y < y_0),$$

$$\chi(0, y) < 0 \quad (y_0 < y < \tfrac{2}{3}).$$

Consequently, along the line $y = k$ $(0 < k < y_0)$, $F^*(x, k)$ is an increasing function of x with $F^*(0, k) < 0$ and $F^*(1, k) > 0$, so that $F^*(x, k)$ vanishes for one and only one value of x, and, for this value of x, $\partial F^*(x, k)/\partial x$ is positive.

Next, along the line $y = k$ $(y_0 < k < \frac{2}{3})$, $\chi(x, k)$ cannot be negative all the way; for, if it were, $F^*(x, k)$ would be a decreasing function of x all along the line, whereas

$$F^*(1, k) = 4k > -2 = F^*(0, k).$$

Hence, as x increases from 0 to 1, $\chi(x, k)$ is first negative and subsequently positive; that is to say, $F^*(x, k)$ decreases from -2 to a (negative) minimum and then increases up to the positive value $4k$. It therefore vanishes for one and only one value of x between 0 and 1; and, for this value of x, it is increasing, so that $\partial F^*(x, k)/\partial x$ is positive.

Now the slope of the curve $F^*(x, y) = 0$ is given by the formula

$$\frac{\partial F^*(x, y)}{\partial x} + \frac{\partial F^*(x, y)}{dy}\frac{dy}{\partial x} = 0;$$

and so, since at all points on the curve inside the rectangle R^* both of the partial differential coefficients $\partial F^*(x, y)/\partial x$ and $\partial F^*(x, y)/\partial y$ are positive, the curve must have a (continuous) negative slope. The part of the curve inside the rectangle R^* consequently consists of an arc steadily falling and moving to the right from $(x_2, \frac{2}{3})$ to $(1, 0)$. Apart from the two points just named, no point of the boundary of R^* is a point of the curve.

The shape of the curve $F^*(x, y) = 0$ is shown by a dotted line in Fig. 2.

The University,
Birmingham.

FUNCTION THEORY IN AN AXIOM-FREE EQUATION CALCULUS

By R. L. GOODSTEIN.

[Received 25 June, 1941.—Read 26 June, 1941.]

Introduction.

Though formalism and finitism have long been deemed antithetical tendencies in enquiries concerning the foundations of mathematics, one of the objectives of recent research has been the reconciliation of the claims of formalist and finitist schools of thought by means of the construction of formal calculi in which the *an-sich* conception of infinity finds no expression. The first step in this direction was taken by Th. Skolem† in 1923. He showed that, in the classical number theory, quantifiers, and the concept of the least number with a certain property, need only to be applied to a bounded range of numbers and, further, that these bounded operators can be defined by primitive recursive definitions. In 1931, Godel‡ showed that every recursive formula can be brought into the form "$A = 0$", where "A" is an expression containing only symbols for primitive recursive functions. These ideas were unified and developed by Bernays in the *Grundlagen der Mathematik*§, where it is proved that the classical number theory can be formalised in a free variable calculus by means of the propositional calculus (extended to cover the rules of substitution for free variables), equality axioms, the axiom of mathematical induction, and by the use of primitive recursive, and explicit, function-definitions. The formulae of this calculus are finite series of signs, each sign being either "0", a free variable number sign, the successor-function sign "x'", a recursive function sign, the sign "$=$", or a sign of the propositional

† "Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichen Ausdehnungsbereich", *Vidensk. Skrifter*, 1, *Mat. Natw. Klasse* (Kristiania, 1928), No. 6.

‡ "Über formal unentscheidbare Sätze . . .", *Monatshefte für Math. und Physik*, 38 (1931), 173–198.

§ Hilbert-Bernays, *Grundlagen der Mathematik*, 1 (1934), § 7, 287–343.

calculus. The logical signs may be eliminated from the formulae, though not from the operations, of the calculus by means of the equivalences:

$$(a = b) \equiv \big(\delta(a, b) = 0\big), \quad (a \neq b) \equiv \big(\{1 \dot{-} \delta(a, b)\} = 0\big),$$

$$(a = 0 \vee b = 0) \equiv (a \,.\, b = 0), \quad (a = 0 \,\&\, b = 0) \equiv (a + b = 0),$$

where "$a = 0$" and "$b = 0$" are formulae of the calculus and, for free variables m, n, the functions $m+n$, $m \,.\, n$, $m \dot{-} n$ and $\delta(m, n)$ have the recursive definitions

$$m+0 = m, \quad m+n' = (m+n)'; \quad m \,.\, 0 = 0, \quad m' \,.\, n = m \,.\, n + m;$$

$$d(0) = 0, \quad d(m') = m; \quad m \dot{-} 0 = m, \quad m \dot{-} n' = d(m \dot{-} n)$$

and

$$\delta(m, n) = (m \dot{-} n) + (n \dot{-} m).$$

In the present paper we consider a further development of the free variable calculus in which a complete elimination of the logical signs leads to an axiom-free, self-sustained, pure equation (or transformation) calculus. This calculus serves for the expression of not only the classical number theory but also a wide part of the classical function theory, and affords a new interpretation of the fundamental concepts of mathematics†.

The author acknowledges with gratitude his indebtedness to Prof. P. Bernays for much generous and valuable advice and assistance.

The equation calculus.

The calculus which we are about to describe operates with four principal categories of signs, numeral variables a, b, c, x, y, z, etc., definite numerals, variable function signs $f(x)$, $\phi(x, y)$, $Ff(x)$, etc., and definite function signs. The characteristic of a numeral variable "x" is that it may be replaced by the sign "0" or the definite function sign "x'". Definite numerals are obtained from the sign "x" by eliminating "x", using only the operations of replacing "x" by "0" or by "x'"; thus

† The elimination of the propositional calculus from the formalisation of the recursive number theory is discussed in a paper by H. B. Curry, "Formalisation of recursive arithmetic", *American J. of Math.*, 63 (1941), 263–282.

"0", "0^{I}", "0^{II}" are definite numerals, "0^{II}", for instance, being obtained from "x" by replacing "x" by "x^{I}", repeating the operation, giving "x^{II}" and finally replacing "x" by "0".

There is a use of variable signs in mathematical literature in direct opposition to that which has just been formulated, the use of variable signs in conditional equations. For example, the "solution" of the equation $2x+5=3+4x$ is said to be the equation $x=1$. The "x" in these "equations" may not be replaced by "0" or by "x^{I}" but only by unity. If we begin with the equation "$2.1+5=3+4.1$" and then omit the sign "1" at each point of its occurrence in the equation, replacing it by a dot or an "x", to show that a sign has been omitted, we obtain the expression "$2x+5=3+4x$". Such a *missing-sign expression* has no place in the calculus, but this is not to say that simple equations cannot be solved in the calculus, only that the calculus cannot formulate the "problem" nor say that it has been solved. That which appertains to the calculus in the process of finding the numeral which was replaced by "x" to form the expression "$2x+5=3+4x$" is the set of equations: $4x-2x=2x$, $5-3=2$, $2.1=2$, $4.1-2.1=2.1$, $4.1-2.1=5-3$, $2.1+5=3+4.1$, the last equation presenting the solution of the problem. Each of these equations is *provable* in a sense which we shall consider later, and each appearance of "x" is as a sign which may be replaced by "0" or "x^{I}". This solution differs from the conventional one only in the omission of the initial expression "$2x+5=3+4x$" and the final "$x=1$"; the equation $2.1+5=3+4.1$ *shows* that "1" is the sign which "x" replaced in forming the initial expression, but *that* a missing sign expression is formed by the omission of a certain sign (or signs) at certain positions is not a statement of the calculus but a comment on it and accordingly cannot be *expressed* in the calculus but only *shown*.

Introductory equations.

The introduction of definite function signs is effected by means of *introductory equations* of specified forms. The introduction of definite function signs is progressive. Starting with definite numerals, numeral variables and the function x^{I}, new function signs are introduced by means of function signs already introduced. The familiar elementary functions, the sum function $\mathrm{Sum}(x, y)$, the product function $\mathrm{Prod}(x, y)$ and the exponential function $\mathrm{Exp}(x, y)$ are introduced by *recursive* equations of introduction which have the form

$$F(x, 0)=a(x), \quad F(x, y^{\mathrm{I}})=b\big(x, y, F(x, y)\big).$$

"Sum (x, y)", "Prod (x, y)" and "Exp (x, y)" are introduced by the equations

$$\text{Sum}\,(x, 0) = x, \quad \text{Sum}\,(x, y') = \big(\text{Sum}\,(x, y)\big)';$$

$$\text{Prod}\,(x, 0) = 0, \quad \text{Prod}\,(x, y') = \text{Sum}\,\big(\text{Prod}\,(x, y), x\big);$$

and $$\text{Exp}\,(x, 0) = 0', \quad \text{Exp}\,(x, y') = \text{Prod}\,\big(\text{Exp}\,(x, y), x\big)$$

respectively. By means of these equations the function signs are transformed into definite numerals, called the values of the function, when the variable signs are replaced by definite numerals; *e.g.* Prod $(0''', 0'')$ transforms successively into

$$\text{Sum}\,\big(\text{Prod}\,(0''', 0'), 0'''\big), \quad \big(\text{Sum}\,\big(\text{Prod}\,(0''', 0'\big), 0''\big)',$$

$$\big(\text{Sum}\,\big(\text{Prod}\,(0''', 0'), 0'\big)\big)'', \quad \big(\text{Sum}\,\big(\text{Prod}\,(0''', 0'), 0\big)\big)''',$$

$$\big(\text{Prod}\,(0''', 0')\big)''', \quad \big(\text{Sum}\,\big(\text{Prod}\,(0''', 0), 0'''\big)\big)''', \quad \big(\text{Sum}\,(0, 0''')\big)''',$$

$$\big(\text{Sum}\,(0, 0'')\big)'''', \quad \big(\text{Sum}\,(0, 0')\big)''''', \quad \big(\text{Sum}\,(0, 0)\big)''''''', \quad 0'''''''.$$

The sign obtained from the sign "x'" by replacing "x" by a definite numeral is called the successor of that numeral.

Of greater complexity than recursive introductory equations, but of fundamentally similar character, are *multiply-recursive* introductory equations, of which an important type is exemplified by

$$F(x, 0) = a(x), \quad F(0, y') = b(y), \quad F(x', y') = c\big(x, y, F(x, y)\big).$$

Multiply recursive introductory equations are also of the (more general) form†

$$F(x, 0) = a(x), \quad F(0, y') = b(y),$$

$$F(x', y') = c\Big(x, y, F\big(x, d\big(x, y, F(x', y)\big)\big)\Big).$$

† A detailed account of multiple recursion is given in Miss Rosa Péter's papers: "Konstruktion nicht-rekursiver Funktionen" and "Ueber die mehrfache Rekursion", *Math. Annalen*, 111 (1935), 42–60 and 113 (1937), 489–527. W. Ackermann gives an example of a recursion which cannot be reduced to multiple recursion: ["Zur Widerspruchsfreiheit der Zahlentheorie, *Math. Annalen*, 117 (1940), 162–194 and in particular §§ 6–7, 178–184.]

Closed chains.

The second general method of introduction of definite function signs is by means of *closed chains* of function signs. A specific list of function signs (all previously introduced) is said to form a closed chain if the successor of any value of any of the functions of the list is itself a value of one of the functions, "0" being a value of one of the functions, no two functions of the list having a value in common, and no function taking the same value twice.

If the function signs "$d(x)$", "$d(x, y)$" are introduced by the recursive equations

$$d(0) = 0, \quad d(x^1) = x; \quad d(x, 0) = x, \quad d(x, y^1) = d\big(d(x, y)\big),$$

and "$\delta(x, y)$" is introduced by the equation $\delta(x, y) = \mathrm{Sum}\big(d(x, y), d(y, x)\big)$, then the condition that the two functions $f(x)$, $g(x)$ have no common value is expressed by the equation $d\Big(0^1, \delta\big(f(x), g(y)\big)\Big) = 0$.

The condition that a function $f(x)$ does not take the same value twice is expressed by

$$\mathrm{Prod}\Big(\delta(x, y),\ d\Big(0^1,\ \delta\big(f(x), f(y)\big)\Big)\Big) = 0.$$

The introduction of "$\delta(x, y)$" above is an instance of the introduction of a function sign by an *explicit* introductory equation of which the form is $F(x, y) = A\big(p(x, y), q(x, y)\big)$ or one of the analogous forms with more or less variables.

If for a given set of values of the variables a function A takes the value a and a function B takes the value b, then we shall call the pair a, b the value of the pair of functions A, B.

A list of pairs of functions is said to form a *closed chain of pairs* if (i) no pair takes the same value twice, (ii) no two pairs have a common value, (iii) one pair takes the value 0, 0, and (iv) for each pair "A", "B" of the chain, each value of the pair of functions A^1, B and each value of the pair A, B^1 is a value of a function pair of the chain†.

† Closed chains might also be defined as follows. A set of functions forms a closed chain if every value of x is a unique value of a unique function of the set. A set of pairs of functions forms a closed chain of every value of the pair (x, y) is a unique value of a unique pair of functions of the set. This definition is equivalent to that given in the text, and helps to bring out the significance of the chain concept more readily, but it is not so convenient in application.

That the pair of function signs "$f(x, y)$", "$g(x, y)$" have no value in common with the pair "$p(x, y)$", "$q(x, y)$" is expressed by the equation

$$\text{Prod}\Big(d\Big(0', \delta\big(f(x, y), p(u, v)\big)\Big),\ d\Big(0', \delta\big(g(x, y), q(u, v)\big)\Big)\Big) = 0.$$

Closed chains of trios, quartets, etc., may be formulated along analogous lines.

We shall prove subsequently that the function signs (in classical notation) $2x$, $2x+1$ form a closed chain and that the pairs x, x; Sum (x, y'), x; x, Sum (x, y') form a closed chain of pairs. A closed chain expresses the concept "every number takes one, and only one, of such and such forms".

A function sign "$F(x)$" is introduced by means of a closed chain by formulating equations of the form $F\big(c(x)\big) = P(x)$ for each function sign "$c(x)$" of the closed chain: $P(x)$ must be a function sign previously introduced. Similarly a two-variable function sign "$F(x, y)$" is introduced by a closed chain of pairs by formulating equations of the form $F\big(j(x, y), k(x, y)\big) = P(x, y)$ for each function pair "$j(x, y)$", "$k(x, y)$" of the closed chain of pairs, each $P(x, y)$ having been previously introduced†. For example, the function *non-square numbers*, $Nsq(r, s)$, is introduced by the equations

$$Nsq(2r, 0) = (r+1)^2+(2r+1), \quad Nsq(2r, s+1) = (r+s+2)^2+(2r+1),$$

$$Nsq(2r+1, 0) = (r+1)^2+2(r+1), \quad Nsq(2r+1, s+1) = (r+s+2)^2+2(r+1),$$

where the abbreviations "1", "2", "$2x$", "$x+y$", "x^2" stand for "$0'$", "$0''$", "Sum (x, x)", "Sum (x, y)", "Prod (x, x)" respectively. The function signs "s^2", "$Nsq(r, s)$" form a closed chain.

Elimination of the propositional calculus. Definition of proof.

Each formula of the calculus takes the form "$a = b$" and expresses the fact that "a" may replace, or be replaced by, "b", where "a" and "b," are function signs; "b" may be said to be a *transform* of "a".

The object of the calculus is the formulation of *proofs*. A *proof* is a (finite) series of equations each of which is either an introductory equation of a function sign, or an equation of the form "$a = a$" (where "a" is a function sign), or is a *proved* equation. If "$a = b$" is an equation of a proof, then

† When introducing a function sign by a closed chain, in each introductory equation every variable sign which occurs on the right-hand side must occur also on the left.

the equation formed from an equation of a proof by replacing "a" by "b" at one or more places of its occurrence in the equation, is a proved equation. The equation formed by replacing a variable sign, at each place of its occurrence in an equation of a proof, by another variable sign or by a definite numeral or function sign, is a proved equation. If two function signs "a", "b" satisfy the same introductory equations, then "$a = b$" is a proved equation†.

Elimination of induction.

The rule that function signs are equivalent if they satisfy the same introductory equations renders superfluous the postulation of the Principle of Mathematical Induction. Inductive proofs of the equivalence of two functions $f(x)$, $g(x)$ proceed by establishing, first the equation $f(0) = g(0)$, and then the implication "$\big(f(x) = g(x)\big) \rightarrow \big(f(x+1) = g(x+1)\big)$"; the basis of this implication is the expression of $f(x+1)$ as a function of $f(x)$ and $g(x+1)$ as the *same* function of $g(x)$, and so proof by induction consists in showing that two function signs satisfy the same recursive introductory equations‡.

The implication $\big(f(x) = g(x)\big) \rightarrow \big(f(x+1) = g(x+1)\big)$ is expressed in the calculus by the equation

$$\text{Prod}\Big(d\big(1,\ \delta\big(f(x), g(x)\big)\big),\ \delta\big(f(x+1), g(x+1)\big)\Big) = 0. \qquad \text{(i)}$$

By showing that $\phi(x) = 0$ is derivable in the calculus from the equations $\phi(0) = 0$, $\text{Prod}\Big(d\big(1, \phi(x)\big), \phi(x+1)\Big) = 0$, Bernays proved that $f(x) = g(x)$ is derivable from $f(0) = g(0)$ and equation (1) above, so that any classical inductive proof can be translated into proof equations of the calculus.

The proof is as follows. We write§ "$x+y$", "$x \,.\, y$" for "Sum (x, y)",

† In the paper "Eine Bemerkung über die Induktionschemata in der rekursiven Zahlentheorie", *Monatshefte für Math. und Phys.*, 48 (1939), 268–276, Th. Skolem shows that the equality of two functions which satisfy the same recursive introductory equations may be proved by appeal to induction in a *single* variable, even though there are functions, defined by multiple recursion, the definition of which by primitive recursion is impossible.

‡ This connection of induction with recursion has been previously observed by both Wittgenstein and Bernays.

§ It is important to notice that any abbreviation that we use, like "$x+y$" for "Sum (x, y)", is *not* a sign of the calculus and must always be regarded as a temporary expedient which finds no place in a formal proof in the system.

"Prod (x, y)". Suppose that for some expression $p(x)$, which may contain other variables than x, the equations $p(0) = 0$ and $d\big(0', p(x)\big)\, p(x') = 0$, are proved equations. It is required to show that $p(x) = 0$ is a proved equation.

The function sign $a(x)$ is introduced by the equation

$$a(x) = d\big(0', d(0', x)\big),$$

so that $a(0) = 0$, $a(x') = 0'$ are proved equations. Since $a(0) \,.\, 0 = 0$, $a(x') \,.\, x' = x'$, it follows that $a(x) \,.\, x = x$, and since $a(0) + d(0', 0) = 0'$, $a(x') + d(0', x') = 0'$, we have $a(x) + d(0', x) = 0'$. Moreover

$$a(x \,.\, 0) = a(x) \,.\, a(0),$$

$a(0 \,.\, y) = a(0) \,.\, a(y)$ and $a(x' \,.\, y') = a(x') \,.\, a(y')$; proving $a(x \,.\, y) = a(x) \,.\, a(y)$, since "$a(x \,.\, y)$" and "$a(x) \,.\, a(y)$" satisfy the same multiply recursive introductory equations. Hence from a proved equation of the form $A \,.\, B = 0$ we deduce in turn, $a(A \,.\, B) = 0$, $a(A) \,.\, a(B) = 0$, $A \,.\, a(A) \,.\, a(B) = 0$, $A \,.\, a(B) = 0$, $A \,.\, d(0', B) = A$. In particular from $d\big(0', p(x)\big)\, p(x') = 0$ we deduce $d\big(0', p(x)\big) \,.\, d\big(0', p(x')\big) = d\big(0', p(x)\big)$.

Thus if "$fp(x)$" is introduced by

$$fp(0) = 0', \quad fp(x') = fp(x) \,.\, d\big(0', p(x)\big),$$

then $fp(0') = fp(0) \,.\, d\big(0', p(0)\big) = d(0', 0) = 0'$, and

$$fp(x') \,.\, d\big(0', p(x')\big) = fp(x) \,.\, d\big(0', p(x)\big),$$

i.e. $fp(x'') = fp(x')$ and so $fp(x') = 0'$; but $fp(0) = 0'$, therefore $fp(x) = 0'$ and so $d\big(0', p(x)\big) = 0'$. Since $x \,.\, d(0', x) = 0$, $p(x)\, d\big(0', p(x)\big) = 0$ and so $p(x) = 0$.

The equivalence theorem.

Bernays has further observed that to translate a given proof of an equation $A = B$, carried out by means of the propositional calculus, into a proof of the equation in the equation calculus, we require three conditions to be satisfied: (1) that it is possible in the equation calculus to pass from $A = B$ to $\delta(A, B) = 0$; (2) that it is similarly possible to pass from $\delta(A, B) = 0$ to $A = B$; (3) that the translation of the axiom $.\, a = b \,.\, \rightarrow \big(\,.\, f(a) = 0 \,.\, \rightarrow .\, f(b) = 0\,.\big)$, viz., the equation

$$d\big(0', \delta(a, b)\big) \,.\, d\big(0', f(a)\big) \,.\, f(b) = 0,$$

is provable in the calculus. Condition (1) is satisfied by definition, and both conditions (2) and (3) are derivable from the equation

$$\text{Prod}\Big(d\big(0', \delta(x, y)\big), f(x)\Big) = \text{Prod}\Big(d\big(0', \delta(x, y)\big), f(y)\Big), \qquad \text{(ii)}$$

which is proved by showing that

$$\text{Prod}\Big(d\big(0', \delta(m, m)\big), f(m)\Big) = f(m),$$

$$\begin{cases} \text{Prod}\Big(d\Big(0', \delta\big(m, \text{Sum}\,(m, n')\big)\Big), f(m)\Big) = 0, \\ \text{Prod}\Big(d\Big(0', \delta\big(m, \text{Sum}\,(m, n')\big)\Big), f\big(\text{Sum}\,(m, n')\big)\Big) = 0, \end{cases}$$

$$\begin{cases} \text{Prod}\Big(d\Big(0', \delta\big(\text{Sum}\,(m, n'), m\big)\Big), f\big(\text{Sum}\,(m, n')\big)\Big) = 0, \\ \text{Prod}\Big(d\Big(0', \delta\big(\text{Sum}\,(m, n'), m\big)\Big), f(m)\Big) = 0. \end{cases}$$

Since the pairs m, m; m, Sum (m, n'); Sum (m, n'), m form a closed chain, these equations reveal that the functions $\text{Prod}\Big(d\big(0', \delta(m, n)\big), f(m)\Big)$ and $\text{Prod}\Big(d\big(0', \delta(m, n)\big), f(n)\Big)$ satisfy the same introductory equations†. Condition (2) is derived from equation (ii) by replacing "$f(m)$" by "m", giving

$$\text{Prod}\Big(d\big(0', \delta(m, n)\big), m\Big) = \text{Prod}\Big(d\big(0', \delta(m, n)\big), n\Big),$$

by means of which $m = n$ follows from $\delta(m, n) = 0$; and condition (3) is derived from equation (ii), by means of the provable equation

$$\text{Prod}\big(d(0', n), n\big) = 0$$

---

† Bernays has also given a proof of equation (ii) which uses the multiply recursive introductory equations

$$f(a, 0) = A(a), \quad f(0, b') = B(b), \quad f(a', b') = G\big(a, b, f(a, b)\big),$$

instead of the closed chain of pairs m, m; m, Sum (m, n'); Sum (m, n'), m: thus showing that the closed chain concept is redundant as an equalising rule for numeral functions. The proof, in fact, requires only primitive recursion, except for the derivation of the equation $a + d(b, a) = b + d(a, b)$.

(which translates the axiom of excluded middle $\sim p \vee p$), by multiplying both sides of (ii) by $d\big(0', f(x)\big)$.

Elimination of the universal and existential operators.

In a calculus in which numeral variables play the part of blank spaces which may be filled by any numeral, and in which missing-number signs have no role, the universal operator Ax is superfluous, since "$Ax\big(f(x) = g(x)\big)$" purports to say that, in the equation $f(x) = g(x)$, the variable sign "x" may be replaced by any numeral, and this is adequately expressed in the calculus by the equation $f(x) = g(x)$ itself. It is perhaps less obvious that the operator Ax is superfluous in such contexts as

$$\text{“}Ax\big(f(x) = 0 \vee g(x) = 0\big)\text{”}, \quad \text{“}Ax\big(f(x) = 0\big) \vee A(x)\big(g(x) = 0\big)\text{”},$$

since the omission of the operator sign renders these distinct propositions identical. This confusion can be removed without retaining the universal operator, by analysing the disjunctions; the former is expressed in the calculus by "$\text{Prod}\big(f(x), g(x)\big) = 0$" and the latter by

$$\text{“Prod}\big(f(x), g(y)\big) = 0\text{”}.$$

The notion of existence has no part in the calculus.

Inequalities.

The uses of inequality signs are of two kinds, functional and relational. The functional inequalities "numbers greater than or equal to a", "numbers less than or equal to a", are expressed in the calculus by the functions "$\text{Sum}(a, r)$", "$d(a, r)$" respectively, whereas the relational inequalities "a is greater than or equal to b", "b is less than or equal to a" are expressed by the equations

$$\text{“}a = \text{Sum}\big(b, d(a, b)\big)\text{”}, \quad \text{“}b = d\big(a, d(a, b)\big)\text{”},$$

which are mutually derivable. The functional concept "numbers between a and $\text{Sum}(a, r)$" is expressed by the function

$$\text{“Sum}\big(a, d(r, s)\big)\text{”}.$$

The technique of the calculus.

Examples. (1) To prove

$$\text{Sum}(x', y) = \big(\text{Sum}(x, y)\big)'.$$

We have

$$(\alpha)\ \mathrm{Sum}\,(x, 0) = x, \qquad (\beta)\ \mathrm{Sum}\,(x', 0) = x', \qquad (\gamma)\ \big(\mathrm{Sum}\,(x, 0)\big)' = x',$$

$$(\delta)\ \mathrm{Sum}\,(x, y') = \big(\mathrm{Sum}\,(x, y)\big)', \qquad (\epsilon)\ \mathrm{Sum}\,(x', y') = \big(\mathrm{Sum}\,(x', y)\big)',$$

$$(\zeta)\ \big(\mathrm{Sum}\,(x, y')\big)' = \big(\mathrm{Sum}\,(x, y)\big)'', \quad (\eta)\ \mathrm{Sum}\,(x', y) = \big(\mathrm{Sum}\,(x, y)\big)'.$$

Commentary. Equation (β) derives from the introductory equation (α) by substituting "x'" for "x". Equation (γ) is obtained from (α) with the help of the equation $x' = x'$, by replacing the first "x" by "Sum $(x, 0)$", as is permitted by (α). Equation (ϵ) derives from the introductory equation (δ) by the substitution of "x'" for "x", and (ζ) derives from (δ) with the help of the equation $\big(\mathrm{Sum}\,(x, y')\big)' = \big(\mathrm{Sum}\,(x, y')\big)'$. Equations (β), (γ), (ϵ), (ζ) show that "Sum (x', y)" and $\big(\mathrm{Sum}\,(x, y)\big)'$ satisfy the same introductory equations, so that (η) is a proved equation.

(2) To prove

$$\mathrm{Sum}\,(x, y) = \mathrm{Sum}\,(y, x).$$

$$\mathrm{Sum}\,(x, 0) = x, \quad \mathrm{Sum}\,(x', 0) = x', \quad \mathrm{Sum}\,(x', 0) = \big(\mathrm{Sum}\,(x, 0)\big)',$$

$$\mathrm{Sum}\,(0, x') = \big(\mathrm{Sum}\,(0, x)\big)', \quad \mathrm{Sum}\,(0, 0) = \mathrm{Sum}\,(0, 0),$$

$$\mathrm{Sum}\,(0, x) = \mathrm{Sum}\,(x, 0), \quad \mathrm{Sum}\,(y', x) = \big(\mathrm{Sum}\,(y, x)\big)',$$

$$\mathrm{Sum}\,(x, y)' = \big(\mathrm{Sum}\,(x, y)\big)'.$$

The last three equations show that Sum (y, x) and Sum (x, y) satisfy the same introductory equations, which completes the proof.

(3) To prove Prod $(0'', x)$, $\big(\mathrm{Prod}\,(0'', x)\big)'$ forms a closed chain.

$$\mathrm{Prod}\,(0'', x') = \mathrm{Sum}\big(\mathrm{Prod}\,(0'', x), 0''\big), \quad \mathrm{Sum}\,(a, 0'') = \big(\mathrm{Sum}\,(a, 0')\big)',$$

$$\big(\mathrm{Sum}\,(a, 0')\big)' = \big(\mathrm{Sum}\,(a, 0)\big)'', \quad \big(\mathrm{Sum}\,(a, 0)\big)'' = a'',$$

$$\mathrm{Sum}\,(a, 0'') = a'', \quad \mathrm{Sum}\big(\mathrm{Prod}\,(0'', x), 0''\big) = \big(\mathrm{Prod}\,(0'', x)\big)'',$$

$$(\alpha)\ \big(\mathrm{Prod}\,(0'', x)\big)'' = \mathrm{Prod}\,(0'', x'), \qquad (\beta)\ \big(\mathrm{Prod}\,(0'', x)\big)' = \big(\mathrm{Prod}\,(0'', x)\big)'.$$

Equations (α), (β) show that the successor of any value of $\big(\text{Prod}\,(0'', x)\big)'$ is a value of $\text{Prod}\,(0'', x)$ and *vice versa*. The equation $\text{Prod}\,(0'', 0) = 0$ shows that zero is a value of one of the functions; it remains to show that the functions have no common value. For this purpose we shall take for granted the proofs of the principal properties of $d(x, y)$, $\delta(x, y)$, such as

$$d\big(\text{Sum}\,(x, y), y\big) = x, \quad d\big(x, \text{Sum}\,(x, y)\big) = 0, \quad \delta(x, y) = \delta(y, x),$$

$$\delta(x, 0) = x, \quad \delta\big(\text{Sum}\,(x, z), \text{Sum}\,(y, z)\big) = \delta(x, y),$$

and the important special case of this last equation, $\delta(x', y') = \delta(x, y)$. For brevity, we shall write "1", "2" for "0'", "0''" and "xy" for "Prod (x, y)".

We prove $$d\big(1, \delta\big(2x, (2y)'\big)\big) = 0.$$

$$\delta\big(0, (2y)'\big) = (2y)', \quad \delta(0, 1) = 1, \quad \delta(2x', 1) = \delta\big((2x)'', 0'\big),$$

$$\delta\big((2x)'', 0'\big) = \delta\big((2x)', 0\big), \quad \delta\big((2x)', 0\big) = (2x)', \quad \delta(2x', 1) = (2x)',$$

$$\delta\big(2x', (2y')'\big) = \delta\big((2x)'', (2y)'''\big), \quad \delta\big((2x)'', (2y)'''\big) = \delta\big(2x, (2y)'\big).$$

And so $$d\big(1, \delta\big(2 \,.\, 0, (2y)'\big)\big) = 0. \qquad (\alpha)$$

$$d\big(1, \delta(2 \,.\, 0, 1)\big) = 0, \quad d\big(1, \delta(2x', 1)\big) = 0, \quad d\big(1, \delta\big(2x, (2 \,.\, 0)'\big)\big) = 0. \qquad (\beta)$$

$$d\big(1, \delta\big(2x', (2y')'\big)\big) = d\big(1, \delta\big(2x, (2y)'\big)\big). \qquad (\gamma)$$

Thus if "$Z(x, y)$" is introduced by $Z(x, y) = 0$, then $Z(0, y) = 0$, $Z(x, 0) = 0$, and $Z(x', y') = Z(x, y)$, so that, from (α), (β), (γ), "$Z(x, y)$" and "$d\big(1, \delta\big(2x, (2y)'\big)\big)$" satisfy the same multiply recursive introductory equations and so $d\big(1, \delta\big(2x, (2y)'\big)\big) = 0$ is a proved equation.

(4) The proof that the pairs a, a; a, Sum (a, b'); Sum (a, b'), a form a closed chain of pairs proceeds on similar lines. We write "$a+b$" for "Sum (a, b)". The greater part of the proof is trivial, since, for instance, the fact that the pair a', a is a value of a function pair of the list is

expressed by the equation $\delta\big(a', (a+0')\big)+\delta(a, a) = 0$. The interesting point is to show that the values of the pair a', $a+b'$ are values of pairs of functions of the list; from the pair of equations

$$\delta(a', a')+\delta(a', a+0') = 0, \quad \delta(a', a')+\delta(a'+b', a+b'') = 0,$$

we derive

$$\{\delta(a', a')+\delta(a', a+b')\}\{\delta(a', a')+\delta(a'+b', a+b')\} = 0,$$

since 0, b' forms a closed chain, and this equation shows that the values of a', $a+b'$ are *either* values of a, a *or* of a, $a+b'$. Such disjunctions are of course characteristic of proofs of closed chains, since the definition requires of a closed chain $p(x)$, $q(x)$, $r(x)$ say, that for some $a(x)$, $b(x)$, $c(x)$ we have

$$\delta\Big(p(x)', p\big(a(x)\big)\Big).\delta\Big(p(x)', q\big(b(x)\big)\Big).\delta\Big(p(x)', r\big(c(x)\big)\Big) = 0.$$

To complete the proof we must show that none of the pairs of functions have a common value, *i.e.* we must prove

$$\Big\{d\big(1, \delta(x, y)\big)\Big\}\Big\{d\big(1, \delta(x, y+z')\big)\Big\} = 0$$

and two similar equations. We observe that $d\big(1, \delta(0, 0)\big)\, d\big(1, \delta(0, z')\big) = 0$ and $d\big(1, \delta(x', 0)\big)\, d\big(1, \delta(x', z')\big) = 0$, which prove

$$d\big(1, \delta(x, 0)\big)\, d\big(1, \delta(x, z')\big) = 0.$$

Also $$d\big(1, \delta(0, y)\big)\, d\big(1, \delta(0, y+z')\big) = 0,$$

and

$$d\big(1, \delta(x', y')\big)\, d\big(1, \delta(x', y'+z')\big) = d\big(1, \delta(x, y)\big)\, d\big(1, \delta(x, y+z')\big).$$

The last three equations show that $d\big(1, \delta(x, y)\big)\, d\big(1, \delta(x, y+z')\big)$ satisfies the same multiply recursive introductory equations as $Z(x, y)$, proving the equation $d\big(1, \delta(x, y)\big)\, d\big(1, \delta(x, y+z')\big) = 0$.

Some observations on the operation of the free variable calculus.

For functions $f(x)$ of the calculus, the Axiom of Excluded Middle, "$f(x) = 0$ or $f(x) \neq 0$" is expressed by the provable equation

$$f(x)\, d\big(1, f(x)\big) = 0.$$

It is, in fact, readily seen that all the axioms of the propositional calculus are expressed in the free variable calculus by *proved* equations. The equation $d\Big(1, d(1, x)+x\Big) = 0$, which translates $\sim(p \,\&\, \sim p)$, plays the same part in the free variable calculus as the latter rule plays in the propositional calculus. For example, the derivation of $f(x) = 0$ from $.x = 0 . \rightarrow . f(x) = 0 .$, $. x > 0 . \rightarrow . f(x) = 0 .$, is translated by $d(1, x) f(x) = 0$, $xf(x) = 0$, $\Big(x+d(1, x)\Big) f(x) = 0$, $\Big(d(x)+1\Big) f(x) = 0$ (since

$$x+d(1, x) = d(x)+1)\Big),$$

$d(x)f(x)+f(x) = 0$, $d\Big(d(x) f(x)+f(x), d(x)f(x)\Big) = 0$, $f(x) = 0$.

Similarly the derivation of $. f(x) = 0 . \rightarrow . h(x) = 0$, from

$$. f(x) = 0 . \rightarrow . g(x) = 0 . \quad \text{and} \quad . g(x) = 0 . \rightarrow . h(x) = 0 .,$$

is translated into the free variable calculus by means of the equation $d\Big(1, d(1, x) y+d(1, y) z\Big) d(1, x) z = 0$, by replacing x, y, z by $f(x)$, $g(x)$, $h(x)$ respectively, for the three equations

$$d\Big(1, f(x)\Big) g(x) = 0, \quad d\Big(1, g(x)\Big) h(x) = 0,$$

$$d\Big(1, d\Big(1, f(x)\Big) g(x)+d\Big(1, g(x)\Big) h(x)\Big) d\Big(1, f(x)\Big) h(x) = 0$$

give $d\Big(1, f(x)\Big) h(x) = 0$.

Of particular importance is the expression in the calculus of the concept "the least x, less than y, for which $f(x) = 0$". We define $\beta(x) = d(0', x)$ and $\min_f(x, a)$ by the equations†

$$\min_f(0, a) = 0, \quad \min_f(x', a) = \min_f(x, a)+x' \beta\Big(\min_f(x, a)+f(x', a)\Big).$$

The characteristic properties of this function are expressed by the provable equations

$$\beta\Big(f(x, a)\Big) . f\Big(\min_f(x, a), a\Big) = 0, \tag{i}$$

$$y . \beta\Big(f(y, a)\Big) . d(x', y) . d\Big(\min_f(x, a), y\Big) = 0. \tag{ii}$$

† If $f(X, a) = 0$, then $\min_f(X, a)$ is the least value of x, less than X, for which $f(x, a) = 0$.

If, for some function $\phi(a)$, $f\big(\phi(a), a\big) = 0$ is a proved equation, $\phi(a)$ being necessarily greater than zero, we derive from (i) that

$$f\Big(\min_f\big(\phi(a), a\big), a\Big) = 0.$$

By means of (ii) we can show that if $f(0, a) > 0$ and $f\big(\phi(a), a\big) = 0$, then $f(x, a) > 0$ for any $x < \min_f\big(\phi(a), a\big)$. First we show that

$$\min_f\big(\phi(a), a\big) \geqslant 1.$$

From $f(0, a) > 0$ we derive $\beta(x) \,.\, \beta\big(f(x, a)\big) = 0$, and since

$$f\Big(\min_f\big(\phi(a), a\big), a\Big) = 0,$$

then $\beta\Big(\min_f\big(\phi(a), a\big)\Big) = 0$. It is readily seen that $x\delta\big(x, 1+d(x, 1)\big) = 0$, and so, writing $\min_f\big(\phi(a), a\big) = \psi(a)$, we have

$$\beta\big(\psi(a)\big)\,\delta\Big(\psi(a), 1+d\big(\psi(a), 1\big)\Big) = 0, \quad \psi(a)\,\delta\Big(\psi(a), 1+d\big(\psi(a), 1\big)\Big) = 0;$$

since $\psi(a)+\beta\big(\psi(a)\big) = d\big(\psi(a)\big)+1$, it follows that

$$\delta\Big(\psi(a), 1+d\big(\psi(a), 1\big)\Big) = 0,$$

and therefore $\psi(a) = 1+d\big(\psi(a), 1\big)$, *i.e.* $\psi(a) \geqslant 1$.

To complete the proof we require the following two simple results. Observe first that from an equation of the form $p' \,.\, q = 0$ we derive $q = 0$; for

$$q = d(p \,.\, q+q, p \,.\, q) = d(p' \,.\, q, p \,.\, q) = d(0, p \,.\, q) = 0.$$

Furthermore $\phi(a) \geqslant \psi(a)$, for, taking $\phi(a)$ for both x and y in (ii), we find $\phi(a) \,.\, d\big(\psi(a), \phi(a)\big) = 0$ and so, since $\phi(a) > 0$, $d\big(\psi(a), \phi(a)\big) = 0$; but $\beta\big(d(x, y)\big) \,.\, \delta\big(y, x+d(y, x)\big) = 0$ and therefore from $d\big(\psi(a), \phi(a)\big) = 0$ we deduce $\delta\Big(\phi(a), \psi(a)+d\big(\phi(a), \psi(a)\big)\Big) = 0$, whence

$$\phi(a) = \psi(a)+d\big(\phi(a), \psi(a)\big),$$

i.e. $\phi(a) \geqslant \psi(a)$. Now $\psi(a) \geqslant 1$; if $\psi(a) > 1$ we can write $\psi(a) = \lambda''$, and, taking $d(\lambda, \mu)'$ for y and $\phi(a)$ for x in equation (ii), from $\psi(a) = \lambda'' > d(\lambda, \mu)'$

and $\phi(a) \geqslant \psi(a)$, we deduce $\phi(a)^{\mathrm{I}} > d(\lambda, \mu)^{\mathrm{I}}$ and therefore

$$\beta\Big(f\big(d(\lambda, \mu)^{\mathrm{I}}, a\big)\Big) = 0,$$

i.e. $f(x, a) > 0$ for any x less than $\psi(a)$. If $\psi(a) = 1$, since $f(0, a) > 0$, it remains true that $f(x, a) > 0$, for $x < \psi(a)$.

The proofs of equations (i) and (ii) are not difficult but are somewhat lengthy. We outline the proof of (i).

The function $\Phi(x, a)$ is introduced by the equations

$$\Phi(0, a) = f(0, a), \quad \Phi(x^{\mathrm{I}}, a) = \Phi(x, a)\, f(x^{\mathrm{I}}, a).$$

Now

$$\begin{aligned} .\Phi(x^{\mathrm{I}}, a) = 0 .\rightarrow\rightarrow . \big(.f(x^{\mathrm{I}}, a) = 0 .\,\&\,.\Phi(x, a) = 0 .\big) \\ \vee\big(.f(x^{\mathrm{I}}, a) = 0 .\,\&\,.\Phi(x, a) > 0 .\big) \\ \vee\big(.f(x^{\mathrm{I}}, a) > 0 .\,\&\,.\Phi(x, a) = 0 .\big). \quad \text{(A)} \end{aligned}$$

We consider the consequences of the implication

$$.\Phi(x, a) = 0 .\rightarrow\rightarrow .f\Big(\min_f(x, a), a\Big) = 0.$$

(1) If $f(x^{\mathrm{I}}, a) = 0$ and $\Phi(x, a) = 0$, then $f\Big(\min_f(x, a), a\Big) = 0$ and

$$\min_f(x^{\mathrm{I}}, a) = \min_f(x, a) + x^{\mathrm{I}}\beta\Big(\min_f(x, a)\Big).$$

Therefore

$$.\min_f(x, a) = 0 .\rightarrow\rightarrow .f\Big(\min_f(x^{\mathrm{I}}, a), a\Big) = f(x^{\mathrm{I}}, a) = 0 .,$$

$$.\min_f(x, a) > 0 .\rightarrow\rightarrow .f\Big(\min_f(x^{\mathrm{I}}, a), a\Big) = f\Big(\min_f(x, a), a\Big) = 0 .,$$

and so

$$\big(.f(x^{\mathrm{I}}, a) = 0 .\,\&\,.\Phi(x, a) = 0 .\big) \rightarrow\rightarrow .f\Big(\min_f(x^{\mathrm{I}}, a), a\Big) = 0.$$

(2) If $.f(x^{\mathrm{I}}, a) = 0 .\,\&\,.\Phi(x, a) > 0 .$; we show first that

$$.\Phi(x, a) > 0 .\rightarrow\rightarrow .\min_f(x, a) = 0. \qquad (a)$$

The definitions of $\Phi(x, a)$ and $\min_f(x, a)$ show that

$$\begin{aligned} .\Phi(x^{\mathrm{I}}, a) > 0 .\rightarrow\rightarrow \big(.\Phi(x, a) > 0 .\,\&\,.f(x^{\mathrm{I}}, a) > 0 .\big), \\ .f(x^{\mathrm{I}}, a) > 0 .\rightarrow\rightarrow .\min_f(x^{\mathrm{I}}, a) = \min_f(x, a) ., \end{aligned}$$

and so

$$\big(.\Phi(x,a)>0.\twoheadrightarrow.\min_f(x,a)=0.\big)\twoheadrightarrow\big(.\Phi(x^1,a)>0.\twoheadrightarrow.\min_f(x^1,a)=0.\big),$$

proving (a), and therefore

$$\big(.f(x^1,a)=0.\&.\Phi(x,a)>0.\big)$$

$$\twoheadrightarrow\big(.f(x^1,a)=0.\&.\min_f(x,a)=0.\big)\twoheadrightarrow.\min_f(x^1,a)=x^1.$$

$$\twoheadrightarrow.f\big(\min_f(x^1,a),a\big)=f(x^1,a)=0.$$

Thus $$\big(.f(x^1,a)=0.\&.\Phi(x,a)>0.\big)\twoheadrightarrow.f\big(\min_f(x^1,a),a\big)=0.$$

(3) $$\big(.f(x^1,a)>0.\&.\Phi(x,a)=0.\big)\twoheadrightarrow.\min_f(x^1,a)=\min_f(x,a).$$

$$\twoheadrightarrow.f\big(\min_f(x^1,a),a\big)=f\big(\min_f(x,a),a\big)=0.$$

Hence

$$.\Big\{.\Phi(x^1,a)=0.\&\big(.\Phi(x,a)=0.\twoheadrightarrow.f\big(\min_f(x,a),a\big)=0.\big)\Big\}.$$

$$\twoheadrightarrow.f\big(\min_f(x^1,a),a\big)=0.,$$

i.e. $$\big(.\Phi(x,a)=0.\twoheadrightarrow.f\big(\min_f(x,a),a\big)=0.\big)$$

$$\twoheadrightarrow\big(.\Phi(x^1,a)=0.\twoheadrightarrow.f\big(\min f(x^1,a),a\big)=0.\big).$$

Also $$.\Phi(0,a)=f(0,a)=0.\twoheadrightarrow.f\big(\min_f(0,a),a\big)=f(0,a)=0.,$$

and therefore

$$.\Phi(x,a)=0.\twoheadrightarrow.f\big(\min_f(x,a),a\big)=0.$$

But $$.f(x,a)=0.\twoheadrightarrow.\Phi(x,a)=0.$$

and so $$.f(x,a)=0.\twoheadrightarrow.f\big(\min_f(x,a),a\big)=0.$$

This informal proof is readily translated into proof equations of the calculus. For instance, equation (A) above (or rather the general formula of which it is an instance) is translated by the provable equation

$$\beta(ab)(a+b)\big(a+\beta(b)\big)\big(b+\beta(a)\big)=0,$$

and similarly the first implication in section (1) is an instance of the equation

$$\beta(a)\beta(c)\delta\Big(a+b\,.\,\beta(a+c),\,b\Big)=0.$$

To find the least x such that $f(x)\leqslant a$, we introduce $F(x,a)$ by the equation $F(x,a)=d\Big(f(x),a\Big)$, so that the required value of x is the least x for which $F(x,a)=0$. It remains to show that $F(x,a)=0 \rightarrowtail f(x)\leqslant a$; this implication is an instance of the equation

$$\beta\Big(d(x,y)\Big)\,\delta\Big(y,\,x+d(y,x)\Big)=0,$$

which is a consequence of the equations

$$\beta\Big(d(a,a)\Big)\,\delta\Big(a,\,a+d(a,a)\Big)=0,$$

$$\beta\Big(d(a+b^1,a)\Big)\,\delta\Big(a,\,a+b^1+d(a,a+b^1)\Big)=0,$$

and

$$\beta\Big(d(a,a+b^1)\Big)\,\delta\Big(a+b^1,\,a+d(a+b^1,a)\Big)=0,$$

since a,a; $a+b^1,a$; $a,a+b^1$ forms a closed chain of pairs.

Introduction of rational number and function signs.

The sign "p/q" is called a rational number (or fraction) variable sign and the signs formed from "p/q" by replacing "p" and "q" by definite numerals (only non-zero values for q) are definite rational number (fraction) signs. A pair of rational numbers $(p/q)/(r/s)$ transforms into the rational number ps/qr (q, r, s all different from zero). Rational function signs are formed by placing a letter, or group of letters, before brackets containing one or more rational or numeral variable signs, and are transformed into pairs of numeral functions (homogeneous and of the same order) by means of introductory equations. For example, the sum function, Sum $(p/q, r/s)$, is introduced by the equation Sum $(p/q, r/s)=$ Sum $(ps, qr)/qs$; and the product function, Prod $(p/q, r/s)$, by the equation Prod $(p/q, r/s)=pr/qs$.

Note that Sum (ps, qr) and qs are both homogeneous and of order 2. Similarly, pr and qs are both homogeneous and of the same order.

A rational function sign containing a numeral variable "n" is called a sequence (in n)-sign.

The connection between the rational calculus and the numeral calculus is given by the rule that the equation $p/q=r/s$ is a proved equation if, and only if, $ps=qr$ is a proved equation. This rule transforms any proved numeral equation $a=b$ into a proved rational equation $a/1=b/1$.

Inequalities between the rational numbers p/q, r/s are represented by the formulae expressing inequalities between the numerals ps, qr. By means of the function $d(p/q, r/s)$, introduced by the equation

$$d(p/q, r/s) = d(ps, qr)/qs,$$

inequalities may be expressed directly; for instance, from the equation $ps = qr + d(ps, qr)$, expressing $ps \geqslant qr$, we can derive

$$p/q = \mathrm{Sum}\Big(r/s, d(p/q, r/s)\Big).$$

Closed chains of pairs are important also in the introduction of rational function signs; consider, *e.g.*, the introduction of $\mathrm{Osc}\,(p/q)$ by the equations $\mathrm{Osc}\,(a/a) = 1$, $\mathrm{Osc}\Big((a^{\mathrm{I}}+b)/a\Big) = b/a^{\mathrm{I}}$, $\mathrm{Osc}\Big(a/(a+b^{\mathrm{I}})\Big) = a/b$.

Rational-function theory.

The following non-formal account of the function theory of the rational calculus is readily translated into proof equations of the calculus by means of the functions $\mathrm{Qt}\,(x, y)$, $\mathrm{Dec}\,(p/q, n)$, $\delta(p/q, r/s)$, which are introduced by the equations

$$\mathrm{Qt}\,(q, 0) = 0, \quad \mathrm{Qt}\Big(a(p+q)^{\mathrm{I}}+q,\ (p+q)^{\mathrm{I}}\Big) = a$$

$\Big($we observe that $q, 0$; $a(p+q)^{\mathrm{I}}+q$, $(p+q)^{\mathrm{I}}$ forms a closed chain of pairs$\Big)$,

$$\mathrm{Dec}\,(p/q, n) = \mathrm{Qt}\,(10^n p, q)/10^n$$

and

$$\delta(p/q, r/s) = \mathrm{Sum}\Big(d(p/q, r/s), d(r/s, p/q)\Big),$$

where "10" stands for "9^{I}", "9" for "8^{I}" and so on, down to "1" for "0^{I}". For brevity we shall write "$|x-y|$" for "$\delta(x, y)$" and "$(p/q)_n$" for "$\mathrm{Dec}\,(p/q, n)$".

Convergence.

Proofs in which the final equation takes the form

$$\Big|f\Big(\phi f(n, x, y)+p, x/y\Big) - f\Big(\phi f(n, x, y), x/y\Big)\Big|_n = 0 \qquad \text{(C)}$$

are called *proofs of convergence*; the sequence $f(n, x/y)$ is said to be proved convergent in n by the proof and the numeral function $\phi f(n, x, y)$ is called the function associated with the sequence in the convergence proof. A sequence which is proved convergent is called a convergent sequence.

In addition to definite convergent sequences we introduce convergent sequence variables "$f(n, x/y)$", with associated functions "$\phi f(n, x, y)$", introduced by equation (C) above, the function variable $\phi f(n, x, y)$ being replaceable by the definite numeral function associated with some definite rational function replacing $f(n, x/y)$, in the proof of the convergence of that rational function.

Relative convergence.

A function $f(m, n)$ such that

$$|f(m^*, n)-f(m, n)|_k = 0,$$

where m, m^* are numbers not smaller than $Mf(k)$, and n is not smaller than $\lambda f(m, m^*, k)$, is said to be *convergent in m, relative to n*†. By means of the function $\max(a, b)$ introduced by the equations

$$\max(a, a) = a, \quad \max(a, a+b') = a+b', \quad \max(a+b', a) = a+b',$$

we can ensure that the function $Mf(k)$ above is steadily increasing, for we can replace $Mf(k)$ by $Mif(k)$, introduced by the equations

$$Mif(0) = Mf(0), \quad Mif(k') = \max\Big(Mif(k), Mf(k')\Big).$$

The reduced sequence.

If $f(m, n)$ is convergent in n, so that $|f(m, n^*)-f(m, n)|_k = 0$, for $n^* > n = Nf(k, m)$, and if $f(m, n)$ is convergent in m, relative to n, then the sequence

$$Rf(n) = f\Big(Mf(n), Nf\Big(n, Mf(n)\Big)\Big),$$

where $Mf(n)$ is given above, is convergent. For if

$$\nu = \max\Big(Nf\Big(n, Mf(n)\Big), Nf\Big(n^*, Mf(n^*)\Big), \lambda f\Big(Mf(n), Mf(n^*), n\Big)\Big),$$

then
$$\Big|f\Big(Mf(n), Nf\Big(n, Mf(n)\Big)\Big)-f\Big(Mf(n), \nu\Big)\Big|_n = 0$$

† Convergence in m *relative* to n of a function $f(m, n)$ does not involve the convergence of $f(m, n)$ for any fixed value of n. The idea of relative convergence is to make $f(m^*, n)$ close to $f(m, n)$, $m^* \geqslant m$, only by taking n sufficiently great. For example, the function m/n is convergent in m relative to n, though m/n is not convergent for any given value of n.

and $$\left| f\Big(Mf(n^*),\ Nf\big(n^*,\ Mf(n^*)\big)\Big) - f\big(Mf(n^*),\ \nu\big)\right|_n = 0;$$

and $$\left| f\big(Mf(n),\ \nu\big) - f\big(Mf(n^*),\ \nu\big)\right|_n = 0,$$

provided that $n^* > n$, whence

$$|\,Rf(p) - Rf(n^1)\,|_n = 0, \quad p \geqslant n^1.$$

The convergent sequence $Rf(n)$ is called the *reduced* sequence of $f(m, n)$.

It can also be shown that $|\,Rf(n^1) - f(x, y)\,|_n = 0$, provided

$$x \geqslant Mf(n^1), \quad y \geqslant \max\Big(Nf\big(n^1,\ Mf(n^1)\big),\ \lambda f\big(x,\ Mf(n^1),\ n^1\big)\Big).$$

The convergence of the reduced sequence is equivalent to Dedekind's theorem. Dedekind asserted that a convergent sequence of real numbers converges to a real number limit; we translate the concept of a convergent sequence of real numbers (completed infinite process) by a function $f(m, n)$, convergent in n, and convergent in m, relative to n, and show how to *derive* from $f(m, n)$ a convergent sequence $Rf(n)$ which (roughly speaking) approaches $f(m, n)$ as closely as we please. We may express this by saying that $Rf(n)$ translates the traditional $\lim_{m\to\infty}\Big(\lim_{n\to\infty} f(m, n)\Big)$.

Continuity.

A function $f(p/q)$ is said to be continuous in p/q, between the rational numbers a and $b > a$, if $|\,f(x^*) - f(x)\,|_n = 0$, where x is any rational between† a and b and x^* any rational between x and $\min\Big(b,\ x + \theta f(a, b, n)\Big)$, $\theta f(p, q, r)$ being a certain rational function‡. If the function "f" contains more than a single variable sign, then these additional variable signs must, in general, be included in the function "θf".

Relative continuity.

A function $f(n, p/q)$ is said to be continuous in p/q, relative to n, between a and b, if there are functions $\theta f(k)$, $Ncf(p/q, r/s, k)$ such that

† The concept "any rational between p/q and $\mathrm{Sum}\,(p/q, r/s)$" is expressed by the function "$\mathrm{Sum}\,(p/q, d(r/s, t/u))$".

‡ The function $\min(a, b)$ is introduced by the equations

$$\min(a, a) = a, \quad \min(a, a + b^1) = a, \quad \min(a + b^1, a) = a,$$

and $$\min(p/q, r/s) = \min(ps, qr)/qs.$$

$|f(n, x^*)-f(n, x)|_k = 0$ for any x^* between x and $\min\big(b, x+\theta f(k)\big)$, and $n > Ncf(x, x^*, k)$†.

Elimination of irrational numbers.

If $f(n, x)$ is convergent in n, and continuous in x, relative to n, and if $\phi(m, x)$ is convergent in m, then $f\big(n, \phi(m, x)\big)$ is convergent in n, and convergent in m, relative to n. For, from the relative continuity of $f(n, x)$ we have $|f(n, x^*)-f(n, x)|_k = 0$ provided that $|x^*-x|_{\theta f(k)} = 0$ and $n \geqslant Ncf(x, x^*, k)$. From the convergence of $\phi(m, x)$,

$$|\phi(m^*, x)-\phi(m, x)|_k = 0,$$

where $m = M\phi(k, x)$, $m^* \geqslant m$, and therefore

$$\big|f\big(n, \phi(m^*, x)\big)-f\big(n, \phi(m, x)\big)\big|_k = 0 \qquad (1)$$

for $m = M\phi\big(\theta f(k), x\big)$, $n \geqslant Nf\big(\phi(m^*, x), \phi(m, x), k\big)$.

From the convergence of $f(n, x)$

$$\big|f\big(n^*, \phi(m, x)\big)-f\big(n, \phi(m, x)\big)\big|_k = 0, \qquad (2)$$

where $n = Nf\big(k, \phi(m, x)\big)$, $n^* \geqslant n$. Equation (2) shows that $f\big(n, \phi(m, x)\big)$ is convergent in n, and equation (1) that $f\big(n, \phi(m, x)\big)$ is convergent in m, relative to n. The reduced sequence of $f\big(n, \phi(m, x)\big)$ translates the traditional $\lim_{n\to\infty} f\Big(n, \lim_{m\to\infty} \phi(m, x)\Big)$. This translation shows that, in the calculus, the values of functions for "irrational" values of the arguments are *constructed*, not defined. **The notion of a function of an irrational variable plays no part at all**‡. Such a function as "e^x" is a function of

† A function $f(n, p/q)$ continuous in p/q *relative* to n is not necessarily continuous for any fixed n. The difference between $f(n, x)$ and $f(n, x^*)$ is small only for large enough values of n.

‡ Contrast this with Brouwer's treatment of the finitist problem. Brouwer defines a function as a correlation between nests of intervals [L. E. J. Brouwer, *Math. Annalen*, 97 (1926–7), 60–61]. We define only correlations between rationals, use continuity (in a rational field) to fill in the gaps, so to speak, and are thus able to dispense with the concept of a totality of real numbers and also with Brouwer's concepts of "Wahlfolge" and "Species".

two variables, the rational variable x and an (unexpressed) numeral variable, and is introduced by the equation

$$e(n, p/q) = E(n, p, q)/\mathrm{Prod}\Big(\mathrm{Fac}(n), \mathrm{Exp}(q, n)\Big),$$

where $\mathrm{Exp}(x, y)$, $\mathrm{Fac}(x)$ and $E(n, p, q)$ are introduced by $\mathrm{Exp}(x, 0) = 0'$, $\mathrm{Exp}(x, y') = \mathrm{Prod}\Big(x, \mathrm{Exp}(x, y)\Big)$, $\mathrm{Fac}(0) = 0'$, $\mathrm{Fac}(x') = \mathrm{Prod}\Big(x', \mathrm{Fac}(x)\Big)$, and $E(0, p, q) = 0'$,

$$E(n', p, q) = \mathrm{Sum}\Big(\mathrm{Prod}\Big(\mathrm{Prod}(q, n'), E(n, p, q)\Big), \mathrm{Exp}(p, n')\Big),$$

respectively. So too "e" is not a number, but the sequence $e(n, 1/1)$. The function "e^{e^x}" is the reduced sequence of the function $e\Big(n, e(m, x)\Big)$, and the "value" of e^x for an irrational value of the argument, like e, is the reduced sequence of the sequence $e\Big(n, e(m, 1)\Big)$.

Introduction of positive and negative integer signs and integer function signs.

The signs "$+$" and "$-$," may be placed before definite or variable numeral signs and definite or variable function signs. The transformation rules for these signs are $++x = +x$, $--x = +x$, $+-x = -x$, $-+x = -x$. If "a" is replaced by a definite numeral, the signs which result from the signs "$+a$", "$-a$" are called positive and negative integers, respectively, and if "a" is replaced by a numeral function the resulting signs are called positive and negative function signs respectively. If to each function sign of a closed chain we prefix the sign "$+$", ("$-$"), we form a closed chain of positive (negative) function signs. Functions of positive or negative integer variable signs are formed by writing one or more positive or negative variable signs in brackets after a group of letters, and are transformable into positive or negative function signs by introductory equations. We admit but one kind of introductory equation, introduction by closed chains of positive or negative functions. Greek letters are used as *integer variables*, which may be replaced either by positive integer or negative integer variables. (Roman letters will also be used as integer variables when the context does not require any emphasis on the distinction between numeral and integer variables.)

Closed chains of integers are formed from the fundamental chain $-x'$, $+y$ by replacing x or y or both by closed chains of numerals. Closed chains of pairs of integers are formed from the fundamental chain of

pairs $+x, +y$; $+x, -y^{\mathrm{I}}$; $-x^{\mathrm{I}}, +y$; $-x^{\mathrm{I}}, -y^{\mathrm{I}}$ by replacing the pair x, y by a closed chain of pairs in any or all of the pairs of the fundamental chain. *Integer functions* are introduced by means of closed chains of integers.

Examples. $\mathrm{Cub}(\xi)$ is introduced by the equations

$$\mathrm{Cub}(+x) = +\mathrm{Exp}(x, 0^{\mathrm{III}}), \quad \mathrm{Cub}(-x^{\mathrm{I}}) = -\mathrm{Exp}(x^{\mathrm{I}}, 0^{\mathrm{III}}),$$

and $\mathrm{Sum}(\xi, \eta)$ by the equations

$$\mathrm{Sum}(+x, +y) = +\mathrm{Sum}(x, y), \quad \mathrm{Sum}(-x^{\mathrm{I}}, -y^{\mathrm{I}}) = -\mathrm{Sum}(x^{\mathrm{I}}, y^{\mathrm{I}}),$$

$$\mathrm{Sum}\left(+\mathrm{Sum}(x, y^{\mathrm{I}}), -y^{\mathrm{I}}\right) = +x, \quad \mathrm{Sum}\left(+y, -\mathrm{Sum}(x^{\mathrm{I}}, y)\right) = -x^{\mathrm{I}},$$

$$\mathrm{Sum}\left(-y^{\mathrm{I}}, +\mathrm{Sum}(x, y^{\mathrm{I}})\right) = +x, \quad \mathrm{Sum}\left(-\mathrm{Sum}(x^{\mathrm{I}}, y), +y\right) = -x^{\mathrm{I}}.$$

$\Delta(\xi, \eta)$, or $\xi - \eta$, is defined by the equations

$$\Delta(+x, -y^{\mathrm{I}}) = +\mathrm{Sum}(x, y^{\mathrm{I}}), \quad \Delta(-x^{\mathrm{I}}, +y) = -\mathrm{Sum}(x^{\mathrm{I}}, y),$$

$$\Delta\left(+\mathrm{Sum}(x, y), +y\right) = +x, \quad \Delta\left(+x, +\mathrm{Sum}(x, y^{\mathrm{I}})\right) = -y^{\mathrm{I}}, \quad \text{etc.}$$

The transformation rules for pairs of integers are:

$$+p/+q = +p/q, \quad +p/-q = -p/q, \quad -p/+q = -p/q, \quad -p/-q = +p/q.$$

The functions $\mathrm{Mod}(\xi)$, $\mathrm{Mod}(\xi/\eta)$ are introduced by the equations

$$\mathrm{Mod}(+p) = +p, \quad \mathrm{Mod}(-p) = +p, \quad \mathrm{Mod}(\xi/\eta) = \mathrm{Mod}(\xi)/\mathrm{Mod}(\eta).$$

We abbreviate "$\mathrm{Mod}(\xi)$" and "$\mathrm{Mod}(\xi/\eta)$" to "$|\xi|$" and "$|\xi/\eta|$" respectively. The function $|\xi - \eta|$ is introduced by the equation

$$|\xi - \eta| = \mathrm{Mod}\left(\Delta(\xi, \eta)\right).$$

The rôle which was played by the function $d(x, y)$ in the numeral calculus is taken in the integer calculus by the function $\tau(\alpha, \beta)$, introduced by the equations

$$\tau\left(+(x+y^{\mathrm{I}}), +x\right) = +y^{\mathrm{I}}, \quad \tau\left(+x, +(x+y^{\mathrm{I}})\right) = 0,$$

$$\tau(+x, +x) = 0, \quad \tau(+x, -y^{\mathrm{I}}) = +(x+y^{\mathrm{I}}), \quad \tau(-x^{\mathrm{I}}, +y) = 0,$$

$$\tau\left(-(x^{\mathrm{I}}+y^{\mathrm{I}}), -x^{\mathrm{I}}\right) = 0, \quad \tau\left(-x^{\mathrm{I}}, -(x^{\mathrm{I}}+y^{\mathrm{I}})\right) = +y^{\mathrm{I}},$$

and $\tau(-x^1, -x^1) = 0$. We can readily prove that

(1) $\tau(a, \beta)+\tau(\beta, a) = |a-\beta|$, (2) $a+\tau(\beta, a) = \beta+\tau(a, \beta)$,

(3) $\tau\big(0, \tau(a, \beta)\big) = 0$, (4) $\tau\Big(\tau(a, \beta)+\tau(\beta, a), \tau(a, \beta)\Big) = \tau(\beta, a)$.

Hence from $|a-\beta| = 0$ we derive $\tau\big(|a-\beta|, \tau(a, \beta)\big) = 0$, by (3), and then by (1) and (4) that $\tau(\beta, a) = 0$ and by (1) again that $\tau(a, \beta) = 0$. Thus from $|a-\beta| = 0$ we derive $\tau(a, \beta) = \tau(\beta, a) = 0$ and thence, by (2), that $a = \beta$.

If $N(a)$ is introduced by the equations $N(0) = 1, N(x^1) = 0, N(-x^1) = 0$, then the equation $N\big(\tau(\beta, a)\big) . \Big|a-\big(\beta+\tau(a, \beta)\big)\Big| = 0$, is provable and serves to derive $a = \beta+\tau(a, \beta)$ from $\tau(\beta, a) = 0$. The inequalities $a \geqslant b$, $a \leqslant b$ are expressed by $a = b+|a-b|$, $a = b-|b-a|$ and the functional inequalities "numbers not less than a", etc., are expressed by $a+|r|$, etc.

The integer calculus.

We shall take for granted the modifications necessary to extend the formulations we have given of convergence and continuity proofs to the integer calculus. We shall, however, consider in some detail the fundamental theorem on continuous functions, that *if $f(x)$ is continuous and takes both positive and negative values, then we can introduce a convergent sequence $s(n)$ and a numeral function $Nfs(n)$ such that $\big|f\big(s(n)\big)\big|_\kappa = 0$ for $n \geqslant Nfs(\kappa)$, i.e. $f\big(s(n)\big)$* can be made as small as we please by sufficiently increasing n.

Proof. Suppose that $f(x)$ is continuous between a and b, and $f(a)$ negative, $f(b)$ positive. Since $f(x)$ is continuous, we can find a steadily increasing numeral function $\phi(\kappa)$ so that the numbers $a_r^\kappa = a+r/2^{\phi(\kappa)}$, between a and b, satisfy

$$|f(a_{r+1}^\kappa)-f(a_r^\kappa)|_\kappa = 0, \tag{1}$$

and since $f(b)$ is positive and $f(a)$ negative, we can introduce a *least* r, r_κ, such that $f(a_{r_\kappa+1}^\kappa)$ is positive and $f(a_{r_\kappa}^\kappa) \leqslant 0$, whence, by equation (1), $|f(a_{r_\kappa}^\kappa)|_\kappa = 0$ and $|f(a_{r_\kappa+1}^\kappa)|_\kappa = 0$, and $|f(x)|_\kappa = 0$ for any x between $a_{r_\kappa}^\kappa$ and $a_{r_\kappa+1}^\kappa$. Again, if $a_{r_\kappa}^\kappa = a_{s_\kappa}^{\kappa+1}$ and $a_{r_\kappa+1}^\kappa = a_{t_\kappa}^{\kappa+1}$, then we can introduce a least r, $r_{\kappa+1}$, between s_κ and t_κ such that $f(a_{r_{\kappa+1}+1}^{\kappa+1})$ is positive and $f(a_{r_{\kappa+1}}^{\kappa+1}) \leqslant 0$, so that $|f(a_{r_{\kappa+1}}^{\kappa+1})|_{\kappa+1} = 0$. Since $\phi(\kappa+1) > \phi(\kappa)$,

$$|a_{r+1}^{\kappa+1}-a_r^{\kappa+1}| \leqslant |a_{r+1}^\kappa-a_r^\kappa|/2,$$

and the convergence of the sequence $a_{r_\kappa}^\kappa$ is assured.

To illustrate the way in which the steps of this proof are translated into the formal calculus, we give the detailed formal proof of the theorem that the equation $x^2 = a+1$ has a solution, that is to say, that there is a convergent sequence $s(n, a)$ such that, for each n, $s(n, a)^2$ differs from $a+1$ by less than $1/10^n$.

We introduce the abbreviations "1", "2", ..., "9", "10" by the equations $1 = 0^1$, $2 = 1^1$, ..., $10 = 9^1$ respectively. We shall also write "$a+b$", "$a.b$", "a^b" for "Sum(a, b)", "Prod(a, b)", "Exp(a, b)" and "$a \geqslant b$", "$a \leqslant b$" for the *equations*

$$\text{“}a = b+d(a, b)\text{”}, \quad \text{“}a = d\big(b, d(b, a)\big)\text{”}$$

respectively, and "$a > b$", "$a < b$" for "$a \geqslant b+1$", "$a+1 \leqslant b$".

Proof. (1) $\lambda(x, a) = d(a, x^2)$; (2) $(a+1)^2 = a+1+a(a+1)$;

(3) $(a+1)^2 \geqslant a+1$; (4) $\big(10^r(a+1)\big)^2 \geqslant 10^{2r}(a+1)$;

(5) $\lambda\big(10^r(a+1),\ 10^{2r}(a+1)\big) = 0$;

(6) $\lambda\Big(\min_\lambda\big(10^r(a+1),\ 10^{2r}(a+1)\big),\ 10^{2r}(a+1)\Big) = 0$;

(7) $10^{2r}(a+1) \geqslant 1$; (8) $\lambda\big(0,\ 10^{2r}(a+1)\big) \geqslant 1$;

(9) $\min_\lambda\big(10^r(a+1),\ 10^{2r}(a+1)\big) \geqslant 1$.

(10) $\theta(a, r) = d\Big(\min_\lambda\big(10^r(a+1),\ 10^{2r}(a+1)\big)\Big)$;

(11) $\min_\lambda\big(10^r(a+1),\ 10^{2r}(a+1)\big) = \theta(a, r)+1$;

(12) $\lambda\big(\theta(a, r)+1,\ 10^{2r}(a+1)\big) = 0$;

(13) $\lambda\big(10\theta(a, r)+10,\ 10^{2(r+1)}(a+1)\big) = 0$;

(14) $\theta(a, r+1) \leqslant 10\theta(a, r)+10$; (15) $\lambda\Big(d\big(\theta(a, r), s\big),\ 10^{2r}(a+1)\Big) \geqslant 1$;

(16) $\lambda\Big(d\big(10\theta(a, r), s\big),\ 10^{2(r+1)}(a+1)\Big) \geqslant 1$; (17) $\theta(a, r+1) \geqslant 10\theta(a, r)$.

(18) $\gamma(a, r) = d\big(10\theta(a, r)+10,\ \theta(a, r+1)\big)$;

(19) $\beta(a, r) = d\big(10,\ \gamma(a, r)\big)$; (20) $\theta(a, r+1) = 10\theta(a, r)+\beta(a, r)$.

(21) $\sigma(a, r, o) = \beta(a, r)$; (22) $\sigma(a, r, p+1) = 10\sigma(a, r, p)+\beta(a, r+p+1)$.

(23) $10\theta(a, r)+\sigma(a, r, 0) = 10\theta(a, r)+\beta(a, r)$;

(24) $\theta(a, r+p+2) = 10\theta(a, r+p+1)+\beta(a, r+p+1)$;

(25) $10^{p+2}\theta(a, r)+\sigma(a, r, p+1)$

$$= 10\Big(10^{p+1}\theta(a, r)+\sigma(a, r, p)\Big)+\beta(a, r+p+1);$$

(26) $\theta(a, r+p+1) = 10^{p+1}\theta(a, r)+\sigma(a, r, p)$;

(27) $|\theta(a, r+p+1)-10^{p+1}\theta(a, r)| = +\sigma(a, r, p)$.

(28) $\tau(a, r, o) = 10$; (29) $\tau(a, r, p+1) = 10\tau(a, r, p)+10$;

(30) $\rho(a, r, 0) = d\Big(10, \beta(a, r)\Big)$;

(31) $\rho(a, r, p+1) = 10\rho(a, r, p)+d\Big(10, \beta(a, r+p+1)\Big)$;

(32) $\mu(a, r, p) = \sigma(a, r, p)+\rho(a, r, p)$;

(33) $\mu(a, r, 0) = \beta(a, r)+d\Big(10, \beta(a, r)\Big)$;

(34) $\beta(a, r)+d\Big(10, \beta(a, r)\Big) = 10$; (35) $\mu(a, r, 0) = 10$;

(36) $\sigma(a, r, p+1)+\rho(a, r, p+1) = 10\Big(\sigma(a, r, p)+\rho(a, r, p)\Big)+10$;

(37) $\mu(a, r, p+1) = 10\mu(a, r, p)+10$;

(38) $\mu(a, r, p) = \tau(a, r, p)$; (39) $\sigma(a, r, p) \leqslant \tau(a, r, p)$.

(40) $\tau(a, r, o) = 10$; (41) $\tau(a, r, 1) = 10\tau(a, r, 0)+10$;

(42) $\tau(a, r, 1) = 10^2+10$; (43) $\tau(a, r, 1) = 10^2+\tau(a, r, 0)$;

(44) $\tau(a, r, p+1)+10^{p+3} = 10\Big(\tau(a, r, p)+10^{p+2}\Big)+10$;

(45) $\tau(a, r, p+2) = 10\Big(\tau(a, r, p+1)\Big)+10$;

(46) $\tau(a, r, p+1) = \tau(a, r, p)+10^{p+2}$;

(47) $10\tau(a, r, p)+10 = \tau(a, r, p)+10^{p+2}$;

(48) $9\tau(a, r, p)+10 = 10^{p+2}$; (49) $\tau(a, r, p) \leqslant 10^{p+2}$;

(50) $\sigma(a, r, p) \leqslant 10^{p+2}$; (51) $10^n\sigma(a, r, p) < 10^{p+n+3}$;

(52) $10^n\sigma(a, n+2, p) < 10^{p+n+3}$;

(53) $\mathrm{Qt}\left(10^n \sigma(a, n+2, p),\ 10^{p+n+3}\right) = 0;$

(54) $\left|\theta\left(a,\ (n+2)+p+1\right) \Big/ 10^{(n+2)+p+1} - \theta(a, n+2) \Big/ 10^{n+2}\right|_n = 0.$

(55) $\theta(a, r) \leqslant 10^r(a+1);$ (56) $2\theta(a, r)+1 \leqslant 10^{a+r+2};$

(57) $\delta\left((x+1)^2,\ x^2\right) = 2x+1;$ (58) $\delta\left(\left(\theta(a, r)+1\right)^2,\ \theta(a, r)^2\right) \leqslant 10^{a+r+2};$

(59) $\left(\theta(a, r)+1\right)^2 > 10^{2r}(a+1);$ (60) $\left(\theta(a, r)\right)^2 \leqslant 10^{2r}(a+1);$

(61) $\delta\left(\left(\theta(a, r)+1\right)^2,\ 10^{2r}(a+1)\right) + \delta\left(10^{2r}(a+1),\ \left(\theta(a, r)\right)^2\right)$

$$= \delta\left(\left(\theta(a, r)+1\right)^2,\ \left(\theta(a, r)\right)^2\right);$$

(62) $\delta\left(10^{2r}(a+1),\ \left(\theta(a, r)\right)^2\right) \leqslant 10^{a+r+2};$

(63) $\left|10^{2r}(a+1) - \left(\theta(a, r)\right)^2\right| \leqslant 10^{a+r+2};$

(64) $\mathrm{Qt}\left(10^n \left(\left|10^{2(a+n+3)}(a+1) - \left(\theta(a, a+n+3)\right)^2\right|\right),\ 10^{2(a+n+3)}\right) = 0;$

(65) $\left|(a+1) - \left(\theta(a, a+n+3)/10^{a+n+3}\right)^2\right|_n = 0.$

Commentary. Equation (1) introduces the function $\lambda(x, a)$. Proof of equation (2) is trivial and is omitted. (3) is derivable from (2) by means of the equation $d\left((a+1)^2,\ a+1\right) = a(a+1)$ and (4) follows from (3). Equation (5) is derivable from (4) and the equation $d(p, p+q) = 0$. Equation (6) expresses the fundamental property of the function

$$\min_\lambda \left(x,\ 10^{2r}(a+1)\right),$$

and is provable by means of (5), as we have shown in the foregoing account of the general function $\min_f(x, a)$. Proof of (7) is omitted, (8) follows from (7), and (9) is derived from (8) as in the general theory of $\min_f(x, a)$. Equation (10) introduces $\theta(a, r)$ and (11) follows from (9); (13) is easily derived from (12), and (14) is a consequence of (13) and the general properties of $\min_f(x, a)$; (17) follows similarly from (16). (18) and (19) are introductory equations, and (20) is a consequence of (17). Equations (20), (23), (24) and (25) show that the functions $\theta(a, r+p+1)$ and

$10^{p+1}\theta(a, r)+\sigma(a, r, p)$ satisfy the same introductory equations, thus proving (26). (27) follows from (26) and the definition of $|x-y|$. (28)–(32) are introductory equations. (34) is a consequence of (19) which says that $\beta(a, r) \leqslant 10$. Equations (35), (37) show that $\mu(a, r, p)$ satisfies the introductory equations of $\tau(a, r, p)$, viz. (28), (29), and so proves (38). Equation (39) follows from (38) and (32). Equations (40)–(47) present a straightforward proof of (48). (53) follows from (52) and the fact that $\mathrm{Qt}(x, y) = 0$ if $x < y$ (*i.e.* $x+1 \leqslant y$). Equation (54) proves that the sequence $\theta(a, r)/10^r$ is convergent (in r). Proofs of (55), (56) are trivial and are omitted. (59) and (60) are also consequences of the fundamental properties of $\min_f(x, a)$. (61) follows from (59) and (60). (63) is just (62) in a different notation. Equation (65) proves that the sequence $\left(\theta(a, r)/10^r\right)^2$ converges (in r) to $a+1$.

The fundamental theorem for relative continuity.

The fundamental theorem, that a continuous rational function cannot change sign without taking the value zero, cannot readily be extended to cover the case of a function $f(n, x)$, convergent in n, and continuous in x, relative to n. The most that we can show, without additional information about $f(n, x)$, is that, if $f(n, x)$ changes sign, we can introduce a pair of sequences $s(n)$, $s^*(n)$, and a numerical function $Nfs(n, x)$, such that $|f(n, x)|_\kappa = 0$, for any x between $s(n)$ and $s^*(n)$, and for $n \geqslant Nfs(\kappa, x)$.

Proof. Since $f(n, x)$ is convergent, there is associated with it a function $N(m, x)$ such that $\left|f(n, x)-f\left(N(\kappa, x), x\right)\right|_\kappa = 0$, $n \geqslant N(\kappa, x)$, and $N(\kappa, x) < N(\kappa+1, x)$. We introduce the function $fo(x, \kappa)$ by the equation $fo(x, \kappa) = f\left(N(\kappa+1, x), x\right)$ and observe that, if $fo(x, \kappa) \geqslant 1/10^\kappa$, then $fo(x, \kappa+1) \geqslant 1/10^\kappa - 1/10^{\kappa+1} > 1/10^{\kappa+1}$, and similarly that if

$$fo(x, \kappa) \leqslant -1/10^\kappa,$$

then $fo(x, \kappa+1) < -1/10^{\kappa+1}$.

The continuity, relative to n, of $f(n, x)$, between a and b, permits the introduction of numerical functions $Nf(x, y, \kappa)$ and $10^{\phi f(\kappa)} = \theta f(\kappa)$ such that, if $a^\kappa(r) = a+r\left(|b-a|/\theta f(\kappa)\right)$, then $\left|f(n, x)-f\left(n, a^\kappa(r)\right)\right|_{\kappa+1} = 0$, provided $n \geqslant Nf\left(a^\kappa(r), x, \kappa+1\right)$, x lies between $a^\kappa(r)$, $a^\kappa(r+1)$ inclusive,

and r is a numeral not greater than $\theta f(\kappa)$. Since therefore each of

$$\left|f\left(n, a^\kappa(r)\right)-f\left(n, a^\kappa(r+1)\right)\right|, \quad \left|f\left(n, a^\kappa(r)\right)-fo\left(a^\kappa(r), \kappa\right)\right|,$$

$$\left|fo\left(a^\kappa(r), \kappa\right)-\left|fo\left(a^\kappa(r), \kappa\right)\right|_{\kappa+1}\right|, \quad \left|f\left(n, a^\kappa(r+1)\right)-fo\left(a^\kappa(r+1), \kappa\right)\right|,$$

$$\left|fo\left(a^\kappa(r+1), \kappa\right)-\left|fo\left(a^\kappa(r+1), \kappa\right)\right|_{\kappa+1}\right|$$

is less than $1/10^{\kappa+1}$, for n not less than

$$\max\left(fN\left(a^\kappa(r), a^\kappa(r+1), \kappa+1\right), N\left(\kappa+1, a^\kappa(r)\right), N\left(\kappa+1, a^\kappa(r+1)\right)\right),$$

it follows that $\left|\left|fo\left(a^\kappa(r+1), \kappa\right)\right|_{\kappa+1}-\left|fo\left(a^\kappa(r), \kappa\right)\right|_{\kappa+1}\right| \leqslant 5/10^{\kappa+1}$, whence

$$\left|\left|fo\left(a^\kappa(r+1), \kappa\right)\right|_{\kappa}-\left|fo\left(a^\kappa(r), \kappa\right)\right|_{\kappa}\right| \leqslant 1/10^{\kappa}$$

and so, if $fov(x, \kappa) = \left(fo(x, \kappa)\right)_\kappa$, the integers $10^\kappa fov\left(a^\kappa(s), \kappa\right)$ are equal or consecutive for consecutive values of s.

We suppose that we can introduce numeral functions $\mu(a, b)$ and $\nu(a, b)$ such that $f(n, a) < -1/10^{\mu(a, b)}$, $f(n, b) > +1/10^{\mu(a, b)}$, for $n \geqslant \nu(a, b)$; from this we can show that $fov(a, \kappa) < -1/10^\kappa$ and $fov(b, \kappa) > +1/10^\kappa$, for $\kappa > \mu(a, b)$. Since $a = a^\kappa(0)$ and $b = a^\kappa\left(\theta f(\kappa)\right)$, we can introduce a least r, $l(\kappa)+1$, such that $fov\left(a^\kappa\left(l(\kappa)+1\right), \kappa\right) \geqslant +1/10^\kappa$, and $fov\left((a^\kappa(s), \kappa\right) < 1/10^\kappa$, for $s \leqslant l(\kappa)$, and therefore each is less than or equal to zero, and a greatest r, $g(\kappa)$ between 0 and $l(\kappa)$ such that $fov\left(a^\kappa\left(g(\kappa)\right), \kappa\right) \leqslant -1/10^\kappa$ (*i.e.* a *least* σ such that $fov\left(a^\kappa\left(d\left(l(\kappa), \sigma\right)\right), \kappa\right) \leqslant -1/10^\kappa$).

Since $fov\left(a^\kappa\left(l(\kappa)\right), \kappa\right) \leqslant 0$, $fov\left(a^\kappa\left(l(\kappa)+1\right), \kappa\right) \geqslant +1/10^\kappa$ and

$$\left|fov\left(a^\kappa\left(l(\kappa)+1\right), \kappa\right)-fov\left(a^\kappa\left(l(\kappa)\right), \kappa\right)\right| \leqslant 1/10^\kappa,$$

it follows that $fov\left(a^\kappa\left(l(\kappa)+1\right), \kappa\right) = 1/10^\kappa$ and $fov\left(a^\kappa\left(l(\kappa)\right), \kappa\right) = 0$;

similarly we see that $fov\Big(a^\kappa\big(g(\kappa)\big), \kappa\Big) = -1/10^\kappa$. From the pair of equations

$$fov\Big(a^\kappa\big(l(\kappa)+1\big), \kappa\Big) = 1/10^\kappa, \quad fov\Big(a^\kappa\big(g(\kappa)\big), \kappa\Big) = -1/10^\kappa$$

it follows that

$$fov\Big(a^\kappa\big(l(\kappa)+1\big), \kappa+1\Big) > 1/10^{\kappa+1} \quad \text{and} \quad fov\Big(a^\kappa\big(g(\kappa)\big), \kappa+1\Big) < -1/10^{\kappa+1}.$$

But $a^\kappa(r) = a^{\kappa+1}\Big(r \operatorname{Exp}\big(10, |\phi f(\kappa+1) - \phi f(\kappa)|\big)\Big)$, and so both $a^{\kappa+1}\big(g(\kappa+1)\big)$ and $a^{\kappa+1}\big(l(\kappa+1)+1\big)$ lie between $a^\kappa\big(g(\kappa)\big)$ and $a^\kappa\big(l(\kappa)+1\big)$. Thus, if $s(\kappa) = a^\kappa\big(g(\kappa)+1\big)$ and $s^*(\kappa) = a^\kappa\big(l(\kappa)\big)$, then $fov(x, \kappa) = 0$ for any x between $s(\kappa)$ and $s^*(\kappa)$, *i.e.* for such values of x, $|f(n, x)|_\kappa = 0$ for $n \geqslant N(\kappa+1, x)$. We have in fact proved more than we enunciated, for in the first place we have shown *that* $fov(x, \kappa)$ *takes every value* $n/10^\kappa$ *between any two of its values, which is the essential property that* $f(n, x)$ *has in common with a rational continuous function*, and secondly that the sequence $s(\kappa)$ is steadily increasing and bounded above by any $s^*(\kappa)$, and $s^*(\kappa)$ is steadily decreasing and bounded below by any $s(\kappa)$.

Classical mathematics asserts that a steadily increasing (or decreasing) bounded sequence is necessarily convergent, but this is *not* true in our equation calculus.

To determine a *convergent* sequence $t(n)$, such that, if $Rtf(n)$ is the reduced sequence of $f\big(n, t(m)\big)$, then $|Rtf(n)|_\kappa = 0$, $n \geqslant Nfs(\kappa)$, we require more information about the function $f(n, x)$ than is given by continuity in x (relative to n). We mention two sufficient (additional) conditions.

(1) If we can introduce a rational function $\lambda f(a, b)$, and a numeral function $\nu f(x, y)$, such that $|f(n, x) - f(n, y)| \geqslant |x-y| \lambda f(a, b)$, provided $n \geqslant \nu f(x, y)$ and x and y lie between a and b, then, from $fov\big(s(\kappa), \kappa\big) = 0$ and $fov\big(s^*(\kappa), \kappa\big) = 0$, we can deduce that $|s(\kappa) - s^*(\kappa)| < 2/\lambda f(a, b)\, 10^\kappa$, which suffices to prove that both $s(\kappa)$ and $s^*(\kappa)$ are convergent, and so, from $fov\big(s(\kappa), \kappa\big) = 0$, that $|Rsf(n)|_\kappa = 0$.

(2) If we can introduce numerical functions $g(x)$, $h(x)$ such that $|f(n, x)| \geqslant 1/10^{g(x)}$ for $n \geqslant h(x)$, then it is determinate whether, for any rational x and sufficiently great n, $f(n, x)$ is less than zero or not less than

zero, and we may carry out the determination of the convergent sequence $t(n)$ exactly as in the case of a rational continuous function.

The differential calculus.

Differentiability proofs have final equations of the form

$$\left|\Big(f(n, x+y)-f(n, x)\Big)\Big/y-\phi(n, x)\right|_\kappa = 0, \qquad \text{(D)}$$

where x and $x+y$ lie between some a and b, y lies between $-1/10^{\theta(a, b, \kappa)}$ and $+1/10^{\theta(a, b, \kappa)}$ and $n \geqslant Nf\phi(x, y, \kappa)$. $\phi(n, x)$ is called the derived function of $f(n, x)$ relative to n. We can prove that $\phi(n, x)$ and $f(n, x)$ are continuous in x, relative to n.

The fundamental theorem of the differential calculus, Rolle's theorem, is provable in the form: If $|f(n, a)|_\kappa = 0$, $|f(n, b)|_\kappa = 0$ for $n \geqslant N(a, b, \kappa)$, and if a *derived* function $\phi(n, x)$ is introduced by the equation (D) above, then we can introduce rational functions $s(n)$ and $t(n)$, where

$$a \leqslant s(n) < t(n) \leqslant b,$$

such that $|\phi(n, x)|_\kappa = 0$ for any x between $s(\kappa)$ and $t(\kappa)$ and $n \geqslant N\phi(x, \kappa)$. We cannot ensure the convergence of $s(n)$ or $t(n)$ without introducing additional restrictions on $f(n, x)$.

Generality in the equation calculus.

We have made frequent usage of function variable signs, like "$f(x)$", and have tacitly supposed that proved equations are obtained when function variable signs are replaced by definite function signs. For example, if we introduce the function $Mf(x)$ by the equations $Mf(0) = f(0)$, $Mf(x') = \max\Big(Mf(x), f(x')\Big)$, it is easily seen that the equation

$$d\Big(f(x), Mf(x)\Big) = 0$$

is provable, for $d\Big(x, \max(x, y)\Big) = 0$, giving $d\Big(f(x'), \max\Big(f(x'), Mf(x)\Big)\Big) = 0$, *i.e.* $d\Big(f(x'), Mf(x')\Big) = 0$, which, together with $d\Big(f(0), Mf(0)\Big) = 0$, gives

$$d\Big(f(x), Mf(x)\Big) = 0. \qquad (1)$$

It is clear from the proof that, throughout, "$f(x)$" may be replaced by a definite function sign, say $\mathrm{Exp}(0^{\mathrm{II}}, x)$, so that the equation obtained from equation (1) by this substitution is a proved equation. Accordingly we introduce the further transformation rule, that the equation formed from

an equation of a proof by replacing a variable function sign by a definite function sign, at each point of its occurrence in the equation, is a proved equation†.

There is a most important distinction to be drawn between the signs "$f(x)$" and "$Mf(x)$" in the foregoing considerations. We have said that "$f(x)$" may be replaced by a definite function sign, and if "$f(x)$" is so replaced, say by Exp $(0'', x)$, then the result of this replacement, in the sign $Mf(x)$, is the sign "M Exp $(0'', x)$", a *definite function* sign whose equations of introduction are the equations which result from the introductory equations of $Mf(x)$. Yet "$Mf(x)$" itself is neither a definite function sign, nor a variable function sign—in the sense in which $f(x)$ is a variable function sign—for "$Mf(x)$" may *not* be replaced by *any* definite function sign, but only by the definite function sign formed by prefixing "M" to the function sign which replaces "$f(x)$" in the introductory equations of $Mf(x)$. The dependence of "$Mf(x)$" upon "$f(x)$" is shown by the form of the sign. To avoid confusion we use only single letters before a pair of brackets, to denote variable function signs. Signs containing two (or three, etc.) letters before brackets (apart from definite function signs) may be called second (or third, etc.) order variable function signs, and introduced by introductory equations containing variable function signs of lesser order. The only introductory equations that we permit are those formed from the introductory equations of definite function signs, by replacing definite by variable function signs. Such an expression as "$F(x) = f\big(g(x)\big)$" is prohibited, and that which one would indicate thereby is expressed by the introductory equation "$Ffg(x) = f\big(g(x)\big)$"‡. This by no means exhausts the uses made of variable function signs. For instance, convergent sequence variable signs are not function variable signs of any order. The calculus of convergent sequence signs is a calculus with the signs "Conv $f(n)$", "Conv $g(n)$", etc., which together with associated signs "Conv $Nf(n)$", "Conv $Ng(n)$" are *introduced* by the

† To ensure that a variable function sign is replaced by a definite function sign of the correct number of arguments, it might be desirable to add a numeral suffix to every function sign, according to the number of its arguments; thus function signs with a single argument place would carry the suffix 1, *e.g.* $f_1(x)$, those of two argument places the suffix 2, like $f_2(x, y)$, and so on, and then we could say that only a definite function sign of the same suffix may replace a variable function sign.

‡ In the introduction of new function signs by explicit definitions the function sign introduced must contain all the variable signs which appear on the right-hand side of the introductory equation. Thus

$$Ffg(x, y) = f\big(g(x), y\big), \quad Ffgh(x, y) = f\big(g(x), h(y)\big), \quad \text{and} \quad Ffgh(x) = f\big(g(x), h(x)\big).$$

convergence equation. "Conv$f(n)$" is replaceable by any definite function sign for which the convergence equation is a proved equation, "Conv$Nf(n)$" being simultaneously replaced by the definite function sign associated with the definite convergent sequence in the convergence proof. We observe that the introductory equation of "Conv$f(n)$" is obtained from a *proved* equation (in this case the final equation of a convergence proof) by replacing a definite by a variable function sign; this is characteristic of the introductory equations of such signs, other instances which we have met with being the introductory equations of continuous function signs "Cont$f(x)$" and differentiable function signs "Deriv$f(x)$".

Generality in the calculus may be exhibited without the use of variable signs, the generality of the theorem showing itself in the generality of the proof, but a calculus which contains no variable signs can only *show* generality, not *express* it.

References.

Hilbert, D. and Bernays, P., *Grundlagen der Mathematik*, I (Berlin, 1934), II (1939).

Skolem, Th., "Begründung der elementaren Arithmetik", *Vidensk. Skrifter*, 1 (*Mat. Natw. Klasse*), Kristiania (1923), No. 6.

————, "Eine Bemerkung über die Induktionsschemata in der rekursiven Zahlentheorie", *Monatshefte für Math. und Phys.*, 48 (1939), 268–276.

Péter, R., "Begriff der rekursiven Funktion", *Math. Annalen*, 110 (1935), 612–632.

————, "Über die mehrfache Rekursion", *Math. Annalen*, 113 (1937), 489–527.

Gödel, K., "Über formal unentscheidbare Sätze", *Monatshefte für Math und Phys.*, 38 (1931), 173–198.

Gentzen, G., "Die Wiederspruchsfreiheit der reinen Zahlentheorie", *Math. Annalen*, 112 (1936), 493–565.

Ackermann, W., "Zur Wiederspruchsfreiheit der Zahlentheorie", *Math. Annalen*, 117 (1940), 162–194.

Brouwer, L. E. J., "Zur Begründung der intuitionistichen Mathematik", *Math. Annalen*, 93 (1925), 244–257; 95 (1926), 453–472; 96 (1927), 451–488.

————, "Über Definitionsbereiche von Funktionen", *Math. Annalen*, 97 (1927), 60–75.

The University,
Reading.

A PROBLEM OF DIOPHANTINE APPROXIMATION IN QUATERNIONS

By KURT MAHLER.

[Received 19 June, 1941.—Read 26 June, 1941.]

In two previous papers† I studied the upper bound of products of two or three linear polynomials in real or complex integral variables. In the present paper I use the same method to prove the following result:

"*If* $\alpha, \beta, \gamma, \delta, \rho, \sigma$ *are constant quaternions such that*

$$\alpha\bar{\alpha}\delta\bar{\delta}+\beta\bar{\beta}\gamma\bar{\gamma}-\alpha\bar{\gamma}\delta\bar{\beta}-\beta\bar{\delta}\gamma\bar{\alpha}=1,$$

then there are two integral quaternions x, y *satisfying*

$$|\alpha x+\beta y+\rho||\gamma x+\delta y+\sigma|\leqslant\tfrac{1}{2}".$$

In so far as the proof is based on special properties of the quaternions, I am indebted to a paper of Speiser‡ and of course to Hurwitz's classical book *Zahlentheorie der Quaternionen*.

In the first chapter, upper bounds for the maximum of $|x|$ in certain sets of quaternions are derived; I mention in particular Theorems 2, 8 and 10, which have some interest in themselves. The second chapter deals chiefly with the reduction of Hermitian forms in quaternions; by combining these results with those of the first chapter, the theorem stated is obtained in the usual way.

I am very much indebted to Prof. Mordell for his help with the manuscript.

† *Journal London Math. Soc.*, 15 (1940), 215–236 and 305–320.

‡ *Journal für Math.*, 167 (1932), 88–97.

CHAPTER I. INEQUALITIES.

1. *Notation.*

Let K be the field of all quaternions

$$x = x_0 + x_1 i_1 + x_2 i_2 + x_3 i_3,$$

where x_0, x_1, x_2, x_3 are arbitrary real numbers, and let

$$\bar{x} = x_0 - x_1 i_1 - x_2 i_2 - x_3 i_3,$$

$$|x| = (x_0^2 + x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}},$$

$$S(x) = x + \bar{x} = 2x_0,$$

$$N(x) = x\bar{x} = |x|^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2$$

be the *conjugate to* x, its *absolute value*, its *trace* and its *norm*. We represent x by a point with rectangular coordinates (x_0, x_1, x_2, x_3) in the four-dimensional Euclidian space K, and call this point also x. Then two points x and y have the *distance* $|x-y|$.

By Hurwitz's definition†, x is an *integral quaternion*, if it can be written in the form

$$x = g_0 j + g_1 i_1 + g_2 i_2 + g_3 i_3,$$

where g_0, g_1, g_2, g_3 are rational integers, and j is the quaternion

$$j = \frac{1 + i_1 + i_2 + i_3}{2}.$$

Hence x is integral if *all numbers* $2x_0, 2x_1, 2x_2, 2x_3$ are *even* or all are *odd* rational integers. The set of all integral quaternions forms a *lattice* L which is generated by the four points j, i_1, i_2, i_3, and which is the four-dimensional analogue to the *centred cube lattice* in ordinary space.

Two points x and y in K are called *congruent*, in symbols $x \equiv y$, if their difference $x-y$ lies in L.

2. *A lemma on linear inequalities.*

The following simple lemma is used repeatedly in this paper:

THEOREM 1. *Suppose that the linear inequalities with real coefficients*

$$l_h(x) = a_{h1}x_1 + \ldots + a_{hn}x_n + a_h \geqslant 0 \quad (h = 1, 2, \ldots, m)$$

† "Zahlentheorie der Quaternionen", *Math. Abhandlungen*, 2 (1933), 303–330, in particular p. 309.

define a bounded set S of points $x = (x_1, \ldots, x_n)$ in n-dimensional Euclidian space. Then the maximum of

$$f(x) = (x_1^2 + \ldots + x_n^2)^{\frac{1}{2}},$$

the distance of x from the origin, is assumed only in points of S in which at least n of the functions $l_h(x)$ vanish simultaneously.

Proof. Since $f(x)$ is continuous and bounded, there is at least one point $x^0 = (x_1^0, \ldots, x_n^0)$ of S in which $f(x^0)$ takes its largest value. If less than n of the numbers $l_h(x)$ vanish, then we can find a positive number ϵ and n numbers $x_1^1, \ldots, x_n^1$ not all zero, such that all points

$$x = (x_1^0 + tx_1^1, \ldots, x_n^0 + tx_n^1) \quad \text{with} \quad -\epsilon \leqslant t \leqslant +\epsilon$$

belong to S. For these points,

$$f(x)^2 = at^2 + \beta t + \gamma = \phi(t)$$

with real coefficients a, β, γ, of which $\frac{1}{2}\phi''(0) = a = f(x^1)^2 > 0$. Hence this function is not a maximum for $t = 0$, contrary to hypothesis.

COROLLARY. *The theorem remains true when some of the inequalities $l_h(x) \geqslant 0$, defining S, are replaced by the equations $l_h(x) = 0$.*

This follows from the last proof.

3. *The sets Δ and $\Delta(g)$.*

Let Δ be the set of all quaternions $x = x_0 + x_1 i_1 + x_2 i_2 + x_3 i_3$ such that

(1) $$|x| \leqslant |x - g| \quad \textit{for every integral quaternion } g.$$

This set has the following properties:

THEOREM 2. *(a) Δ is determined by the linear inequalities*

(2) $$|x_0| \leqslant \tfrac{1}{2}, \quad |x_1| \leqslant \tfrac{1}{2}, \quad |x_2| \leqslant \tfrac{1}{2}, \quad |x_3| \leqslant \tfrac{1}{2}, \quad |x_0| + |x_1| + |x_2| + |x_3| \leqslant 1.$$

(b) For all points of Δ,

(3) $$|x| \leqslant \sqrt{\tfrac{1}{2}},$$

with equality if and only if x is one of the 24 points

(4) $$\frac{\mp 1 \mp i_1}{2}, \quad \frac{\mp 1 \mp i_2}{2}, \quad \frac{\mp 1 \mp i_3}{2}, \quad \frac{\mp i_1 \mp i_2}{2}, \quad \frac{\mp i_1 \mp i_3}{2}, \quad \frac{\mp i_2 \mp i_3}{2}.$$

(c) To every quaternion there is a congruent one in Δ.

Proof. (A) *If x satisfies* (2), *then* (3) *also holds.* The formulae (2) can be written as

$$\mp x_0+\tfrac{1}{2}\geqslant 0, \quad \mp x_1+\tfrac{1}{2}\geqslant 0, \quad \mp x_2+\tfrac{1}{2}\geqslant 0, \quad \mp x_3+\tfrac{1}{2}\geqslant 0,$$

$$\mp x_0\mp x_1\mp x_2\mp x_3+1\geqslant 0.$$

Hence, by Theorem 1, the maximum of $|x|$ is attained in a point in which at least *four* of these inequalities hold in the stronger form with the *sign of equality*. This clearly is possible only in the points (4), and these are all of absolute value $\sqrt{\frac{1}{2}}$.

(B) *The set* (2) *is identical with* Δ. If $g=g_0j+g_1i_1+g_2i_2+g_3i_3$ is an integral quaternion, then the inequality $|x|\leqslant|x-g|$ can be written as

$$\gamma_0x_0+\gamma_1x_1+\gamma_2x_2+\gamma_3x_3\leqslant\Gamma,$$

where

$$\gamma_0=\frac{g_0}{2|g|}, \quad \gamma_1=\frac{g_0+2g_1}{2|g|}, \quad \gamma_2=\frac{g_0+2g_2}{2|g|}, \quad \gamma_3=\frac{g_0+2g_3}{2|g|}, \quad \Gamma=\frac{|g|}{2},$$

and therefore $$\gamma_0^2+\gamma_1^2+\gamma_2^2+\gamma_3^2=1.$$

Hence the distance Γ of the hyperplane

$$\gamma_0x_0+\gamma_1x_1+\gamma_2x_2+\gamma_3x_3=\Gamma$$

from the origin is $\frac{1}{2}$ if g is one of the 24 units

$$\epsilon=\mp 1, \ \mp i_1, \ \mp i_2, \ \mp i_3, \ \frac{\mp 1\mp i_1\mp i_2\mp i_3}{2} \tag{5}$$

of K, and it is greater than or equal to $\sqrt{\frac{1}{2}}$, if $g\neq 0$ is any other integral quaternion. Therefore those inequalities (1) which are not included in (2) are a consequence of the latter.

(C) If x is an arbitrary quaternion, then there is an integral quaternion h such that

$$|x-h|\leqslant|x-g-h| \quad \text{for all integral quaternions } g.$$

Hence $x'=x-h\equiv x$ satisfies (1) and therefore lies in Δ. This completes the proof.

For any quaternion g, let $\Delta(g)$ be the set of all points x for which $x-g$ belongs to Δ; in particular $\Delta(0)=\Delta$. $\Delta(g)$ can also be defined as the set of all points whose distance from g is not larger than that from any

other point congruent to g. If g runs over all elements of L, then the sets $\Delta(g)$ together just fill up the space K without overlapping.

4. *The singular vertices and the transformations of* Δ.

The 24 points (4) on the boundary of Δ are called *singular vertices*, or, briefly, SV; the same name is given to all congruent points. If Σ_0 denotes the special SV

$$\Sigma_0 = \frac{1+i_1}{2},$$

then all SV of Δ given by (4) can be written as

$$\Sigma = \epsilon\Sigma_0,$$

where ϵ denotes the 24 units (5) of K. The 24 transformations of K

$$x \to \epsilon x \tag{6}$$

(*i.e.* the association with the point x of a new point ϵx, or replacing x by ϵx), transform the SV (4) of Δ into one another; they leave all distances and the lattice L invariant, and therefore also the set Δ. More generally, the group of transformations

$$x \to \epsilon x + g, \tag{7}$$

where ϵ is a unit and g an integral quaternion, transforms the lattice L and the set of all SV into themselves, preserves all distances, and changes Δ into $\Delta(g)$.

To every SV, there is a sub-group of order 8 of this group (7) which leaves this point unchanged. For the special SV Σ_0, this group consists of the transformations

$$x - \Sigma_0 \to e(x - \Sigma_0), \quad \text{where} \quad e = \mp 1, \mp i_1, \mp i_2, \mp i_3. \tag{8}$$

By this group, Δ is transformed into the 8 sets

$$\Delta(g), \quad \text{where} \quad g = 0,\ 1,\ i_1,\ 1+i_1,\ \frac{1+i_1 \mp i_2 \mp i_3}{2}, \tag{9}$$

these being the only $\Delta(g)$ which meet in the SV Σ_0.

The 24 SV of Δ can be divided into 3 sets of 8 congruent points, namely

$$\left(\frac{\mp 1 \mp i_1}{2},\ \frac{\mp i_2 \mp i_3}{2}\right), \quad \left(\frac{\mp 1 \mp i_2}{2},\ \frac{\mp i_1 \mp i_3}{2}\right), \quad \left(\frac{\mp 1 \mp i_3}{2},\ \frac{\mp i_1 \mp i_2}{2}\right),$$

and all points of the first set can be written as

$$\Sigma = \Sigma_0 - g, \tag{10}$$

where g is again one of the 8 numbers (9).

Finally, every SV in K is congruent to one of the three SV

$$\frac{1+i_1}{2}, \quad \frac{1+i_2}{2}, \quad \frac{1+i_3}{2},$$

and therefore can be written as

$$\Sigma = \frac{a_0+a_1i_1+a_2i_2+a_3i_3}{2},$$

where the a's are integers of which two are even and two odd.

5. *The sets Δ^* and D.*

Let Δ^* be the sub-set of all those points x of Δ which satisfy the 24 inequalities

$$|x-\Sigma_0| \leqslant |x-\Sigma| \quad \text{for all SV } \Sigma \text{ in (4).} \tag{11}$$

THEOREM 3. *(a) The set Δ^* is determined by the inequalities*

$$\begin{cases} 0 \leqslant x_0 \leqslant \frac{1}{2}, \quad 0 \leqslant x_1 \leqslant \frac{1}{2}, \quad \max(|x_2|, |x_3|) \leqslant \min(x_0, x_1), \\ x_0+x_1+|x_2|+|x_3| \leqslant 1. \end{cases} \tag{12}$$

(b) To every x in Δ there is a unit ϵ of K such that $\epsilon^{-1}x$ lies in Δ^*.

Proof. The inequalities (11) are linear in the coordinates of x; it is easily shown that those belonging to the 8 SV

$$\Sigma = \frac{1\mp i_2}{2}, \quad \frac{1\mp i_3}{2}, \quad \frac{i_1\mp i_2}{2}, \quad \frac{i_1\mp i_3}{2}$$

imply the other 16 inequalities. On writing down these 8 inequalities and using the definition (2) of Δ, the formulae (12) follow at once. If x lies in Δ, then let $\Sigma = \epsilon\Sigma_0$ be the SV nearest to it; then Σ_0 is the SV nearest to $\epsilon^{-1}x$, and therefore $\epsilon^{-1}x$ lies in Δ^*.

We define a further set D, which is to consist of all points x in K such that

$$|x-\Sigma_0| \leqslant |x-\Sigma| \quad \text{for all SV } \Sigma. \tag{13}$$

THEOREM 4. *(a) D is determined by the inequality*

$$\max(|x_0-\tfrac{1}{2}|, |x_1-\tfrac{1}{2}|)+\max(|x_2|, |x_3|) \leqslant \tfrac{1}{2}. \tag{14}$$

(b) The set Δ^* consists of all those points of D which lie in Δ.

Proof. Put

$$x-\Sigma_0=\xi=\xi_0+\xi_1 i_1+\xi_2 i_2+\xi_3 i_3,$$

$$\Sigma-\Sigma_0=a=\frac{a_0+a_1 i_1+a_2 i_2+a_3 i_3}{2}.$$

Then either a is an integral quaternion, or one of the numbers of each pair of coefficients (a_0, a_1) and (a_2, a_3) is an *even* and one an *odd* integer, so that a is a SV.

The point ξ satisfies the inequalities

$$|\xi|\leqslant|\xi-a| \quad \text{for all quaternions } a; \tag{15}$$

or written explicitly,

$$a_0\xi_0+a_1\xi_1+a_2\xi_2+a_3\xi_3\leqslant\frac{a_0^2+a_1^2+a_2^2+a_3^2}{4}. \tag{16}$$

The inequalities corresponding to integral a imply that ξ lies in Δ; hence, by Theorem 2,

$$|\xi_0|\leqslant\tfrac{1}{2},\quad |\xi_1|\leqslant\tfrac{1}{2},\quad |\xi_2|\leqslant\tfrac{1}{2},\quad |\xi_3|\leqslant\tfrac{1}{2},\quad |\xi_0|+|\xi_1|+|\xi_2|+|\xi_3|\leqslant 1.$$

But these formulae are a consequence of the inequalities

$$|\xi_0|+|\xi_2|\leqslant\tfrac{1}{2},\quad |\xi_0|+|\xi_3|\leqslant\tfrac{1}{2},\quad |\xi_1|+|\xi_2|\leqslant\tfrac{1}{2},\quad |\xi_1|+|\xi_3|\leqslant\tfrac{1}{2}, \tag{17}$$

which are the inequalities (15) or (16) if a is one of the 16 SV

$$\frac{\mp 1\mp i_2}{2},\quad \frac{\mp 1\mp i_3}{2},\quad \frac{\mp i_1\mp i_2}{2},\quad \frac{\mp i_1\mp i_3}{2}.$$

Now (17) implies (15) for all quaternions a; in order to prove this, it suffices to assume that a is a SV. Let a_κ and a_λ be the two *even*, and a_μ and a_ν the two odd coefficients of a; we may suppose without loss of generality that

$$a_\kappa\geqslant a_\lambda\geqslant 0,\quad a_\mu\geqslant a_\nu\geqslant 0.$$

Then $|\xi_\kappa|\leqslant\frac{1}{2}$, $|\xi_\lambda|\leqslant\frac{1}{2}$; hence

$$|a_\kappa\xi_\kappa+a_\lambda\xi_\lambda|\leqslant\frac{a_\kappa+a_\lambda}{2}\leqslant\frac{a_\kappa^2+a_\lambda^2}{4},$$

since a_κ and a_λ are even and therefore

$$a_\kappa^2\geqslant 2a_\kappa,\quad a_\lambda^2\geqslant 2a_\lambda.$$

Further $|\xi_\mu| \leqslant \frac{1}{2}$, $|\xi_\mu|+|\xi_\nu| \leqslant \frac{1}{2}$; hence

$$|a_\mu\xi_\mu+a_\nu\xi_\nu| \leqslant (a_\mu-a_\nu)|\xi_\mu|+a_\nu(|\xi_\mu|+|\xi_\nu|)$$
$$\leqslant \tfrac{1}{2}(a_\mu-a_\nu)+\tfrac{1}{2}a_\nu = \frac{a_\mu}{2} \leqslant \frac{a_\mu^2+1}{4} \leqslant \frac{a_\mu^2+a_\nu^2}{4},$$

since a_μ and a_ν are odd and therefore

$$2a_\mu \leqslant a_\mu^2+1, \quad a_\nu^2 \geqslant 1.$$

If we now replace ξ by $x = \xi+\Sigma_0$, then (17) changes into (14). The second assertion of the theorem is evident from (2), (12) and (14).

6. *The sets* $\delta(\tau)$, $\delta(\tau|\Sigma)$, *and* $\delta^*(\tau)$.

Let τ be a number in the interval

$$0 \leqslant \tau \leqslant \tfrac{1}{4}, \tag{18}$$

and $\delta(\tau)$ the set of all points x for which

$$|x_0-\tfrac{1}{2}|+|x_1-\tfrac{1}{2}| \leqslant \tau, \quad |x_2|+|x_3| \leqslant \tau. \tag{19}$$

For $\tau = 0$, $\delta(\tau)$ reduces to the single point Σ_0; for $\tau > 0$, Σ_0 is the centre of the set.

The transformations (8) do not change Σ_0 and are easily seen to leave $\delta(\tau)$ invariant. Hence, if the transformation (7) changes Σ_0 into the new SV Σ, then if x lies in $\delta(\tau)$, the point $\epsilon x+g$ lies in a new set $\delta(\tau|\Sigma)$ of centre Σ. This set depends only on τ and Σ but not on the special transformation (7) which changes Σ_0 into Σ. If, in

$$\Sigma = \frac{a_0+a_1i_1+a_2i_2+a_3i_3}{2},$$

a_κ and a_λ are the two even, and a_μ and a_ν the two odd coefficients, then $\delta(\tau|\Sigma)$ is determined by the inequalities

$$\left|x_\kappa-\frac{a_\kappa}{2}\right|+\left|x_\lambda-\frac{a_\lambda}{2}\right| \leqslant \tau, \quad \left|x_\mu-\frac{a_\mu}{2}\right|+\left|x_\nu-\frac{a_\nu}{2}\right| \leqslant \tau.$$

Since we do not use this result, I omit the simple proof.

For $\tau > 0$, the set $\delta(\tau)$ is bounded by 8 hyperplanes. There are 16 sets of 4 of these hyperplanes which intersect in a point; these 16 points of intersection are easily found to be

$$x = \Sigma_0-2\tau s, \quad \text{where} \quad s = \frac{\mp 1 \mp i_2}{2}, \quad \frac{\mp 1 \mp i_3}{2}, \quad \frac{\mp i_1 \mp i_2}{2}, \quad \frac{\mp i_1 \mp i_3}{2}. \tag{20}$$

They form the four-dimensional analogue to the vertices of a cube in ordinary space.

Now let $\delta^*(\tau)$ be that part of $\delta(\tau)$ which lies in Δ; this new set is determined by the inequalities

$$(21) \qquad x_0 \leqslant \tfrac{1}{2}, \quad x_1 \leqslant \tfrac{1}{2}, \quad x_0+x_1 \geqslant 1-\tau, \quad x_0+x_1+|x_2|+|x_3| \leqslant 1.$$

From these inequalities, we get

$$x_0 \geqslant (1-\tau)-\tfrac{1}{2} = \tfrac{1}{2}-\tau \geqslant \tau \geqslant 0, \quad x_1 \geqslant \tau \geqslant 0,$$

$$|x_0-\tfrac{1}{2}|+|x_1-\tfrac{1}{2}| = 1-x_0-x_1 \leqslant \tau,$$

$$|x_2|+|x_3| \leqslant 1-(x_0+x_1) \leqslant 1-(1-\tau) = \tau,$$

$$|x_2| \leqslant \tau \leqslant \tfrac{1}{2}, \quad |x_3| \leqslant \tau \leqslant \tfrac{1}{2},$$

so that x indeed belongs both to Δ and to $\delta(\tau)$. But the stronger result holds that x is also an element of Δ^*; for from the last inequalities

$$\max(|x_2|, |x_3|) \leqslant \tau \leqslant \min(x_0, x_1).$$

Of the vertices (20) of $\delta(\tau)$, only 8 belong to $\delta^*(\tau)$, namely

$$(22) \qquad x = \Sigma_0 - 2\tau s, \quad \text{where} \quad s = \frac{1 \mp i_2}{2}, \quad \frac{1 \mp i_3}{2}, \quad \frac{i_1 \mp i_2}{2}, \quad \frac{i_1 \mp i_3}{2}.$$

They all lie on the hyperplane

$$(23) \qquad x_0+x_1 = 1-\tau$$

and have the same distance

$$(24) \qquad \rho(\tau) = (2\tau^2-\tau+\tfrac{1}{2})^{\frac{1}{2}}$$

from the origin.

The points (22) are the only points of intersection of (23) with sets of 3 of the hyperplanes

$$x_0 = 0 \text{ or } \tfrac{1}{2}, \quad x_1 = 0 \text{ or } \tfrac{1}{2}, \quad x_0+x_1 \mp x_2 \mp x_3 = 1.$$

Hence, by Theorem 1, all points of the set

$$0 \leqslant x_0 \leqslant \tfrac{1}{2}, \quad 0 \leqslant x_1 \leqslant \tfrac{1}{2}, \quad x_0+x_1+|x_2|+|x_3| \leqslant 1, \quad x_0+x_1 \leqslant 1-\tau$$

satisfy the inequality

$$|x| \leqslant \rho(\tau).$$

Therefore this is also true for all points of Δ^* for which

$$x_0+x_1 \leqslant 1-\tau;$$

and conversely, if x *lies in* Δ^* *and satisfies the inequality*

$$|x| \geqslant \rho(\tau),$$

then x *belongs to the set* $\delta^*(\tau)$, *hence also to* $\delta(\tau)$.

From this result, we deduce

THEOREM 5. *Suppose that the point* x *has a distance not less than* $\rho(\tau)$ *from every lattice point. Then there is a* SV Σ, *such that* x *belongs to* $\delta(\tau|\Sigma)$.

Proof. Let Σ be the SV nearest to x. By a suitable transformation (7), we can change Σ_0 into Σ, and a certain point x_0 in Δ^* into x. The assertion has already been proved for the point x_0; it therefore is true also for x, since the transformation (7) preserves all distances, and only interchanges the domains $\delta(\tau|\Sigma)$.

It is useful to remark that if the equation

$$\rho(\tau)^2 = 2\tau^2-\tau+\tfrac{1}{2}$$

is solved for τ, then, since $0 \leqslant \tau \leqslant \frac{1}{4}$,

$$\tau(\rho) = \frac{1-\sqrt{(8\rho^2-3)}}{4} \quad \text{for} \quad \sqrt{\tfrac{3}{8}} \leqslant \rho \leqslant \sqrt{\tfrac{1}{2}}. \tag{25}$$

Outside this interval for ρ, the value of τ has no meaning for our problem.

7. *An upper bound for the points in* $\delta(\tau|\Sigma)$.

Let Σ be one of the 24 SV of Δ, and x a point in $\delta(\tau|\Sigma)$. If x lies in Δ, then we know that $|x| \leqslant \sqrt{\frac{1}{2}}$. A more general result is given by

THEOREM 6. *If* Σ *is a singular vertex of* Δ, *then*

$$|x| \leqslant (2\tau^2+\tau+\tfrac{1}{2})^{\frac{1}{2}} \tag{26}$$

for all points of $\delta(\tau|\Sigma)$.

Proof. The 24 sets $\delta(\tau|\Sigma)$ are interchanged by the transformations (6), and these preserve all distances. Hence, without loss of generality, we

may suppose x to be a point in $\delta(\tau)$. Then, by Theorem 1, the maximum of $|x|$ is attained in one of the 16 points (20). The 8 points (22) belong to Δ and therefore have distances not greater than $\sqrt{\frac{1}{2}}$ from the origin. The distances from the origin of the remaining 8 points

$$x = \Sigma_0 - 2\tau s, \quad \text{where} \quad s = \frac{-1 \mp i_2}{2}, \quad \frac{-1 \mp i_3}{2}, \quad \frac{-i_1 \mp i_2}{2}, \quad \frac{-i_1 \mp i_3}{2},$$

are all equal to

$$(2\tau^2 + \tau + \tfrac{1}{2})^{\frac{1}{2}},$$

which is not less than $\sqrt{\frac{1}{2}}$, and is the minimum required.

THEOREM 7. *If x lies in $\delta(\tau)$, then the 8 points*

$$x^* = x - g, \quad \textit{where} \quad g = 0,\ 1,\ i_1,\ 1+i_1,\ \frac{1+i_1 \mp i_2 \mp i_3}{2},$$

satisfy the inequality

$$|x^*| \leqslant (2\tau^2 + \tau + \tfrac{1}{2})^{\frac{1}{2}}.$$

Proof. This is a special case of the preceding theorem, since the points x^* belong to the sets $\delta(\tau \mid \Sigma)$, where $\Sigma = \Sigma_0 - g$ are the 8 SV of Δ congruent to Σ_0.

Evidently

$$2\tau^2 + \tau + \tfrac{1}{2} = \frac{1 + 2\rho^2 - \sqrt{(8\rho^2 - 3)}}{2}, \tag{27}$$

where ρ is connected with τ by the formulae (24) and (25).

8. *The set $d(\Lambda)$.*

Let Λ be a number in the interval

$$0 \leqslant \Lambda \leqslant 1, \tag{28}$$

and $d(\Lambda)$ the set of all quaternions satisfying

$$0 \leqslant x_3 \leqslant x_2 \leqslant x_1 \leqslant x_0 \leqslant \tfrac{1}{2}, \quad x_0 + x_1 + x_2 + x_3 \leqslant 1, \quad x_0 + x_1 \leqslant \Lambda. \tag{29}$$

THEOREM 8. *For all points of $d(\Lambda)$,*

$$|x|^2 \leqslant \frac{\Lambda}{2}. \tag{30}$$

Proof. We begin with the remark that if $x+y=s$ and $x \geqq s/2$, where s is a constant, then

$$x^2+y^2=\frac{s^2+(2x-s)^2}{2}$$

is a steadily increasing function of x.

Let x be a point of $\mathrm{d}(\Lambda)$, in which the maximum of $|x|$ is attained. Then the following three statements hold:

(31) $$x_0+x_1=\Lambda.$$

(32) *Either* $x_1=x_2=x_3$ *or* $x_0+x_1+x_2+x_3=1$.

(33) *If* $x_0+x_1+x_2+x_3=1$, *then either* $x_0=\frac{1}{2}$, *or* $x_1=x_2$.

For, if $x_0+x_1<\Lambda$ and also $x_0+x_1+x_2+x_3<1$, then $|x|$ is increased when x_0, x_1, x_2, x_3 are replaced by

$$x_0+\epsilon,\ x_1,\ x_2,\ x_3 \quad \text{for} \quad x_0<\tfrac{1}{2},$$

and by

$$x_0,\ x_1+\epsilon,\ x_2,\ x_3 \quad \text{for} \quad x_0=\tfrac{1}{2},$$

where $\epsilon>0$ is a sufficiently small number; for in the second case $x_1<\Lambda-\frac{1}{2}\leqq x_0$. If, however, $x_0+x_1<\Lambda$ and $x_0+x_1+x_2+x_3=1$, so that $x_2+x_3>1-\Lambda\geqq 0$ and therefore x_2 and x_3 are not both zero, then let ν be the larger index 2 or 3 for which $x_\nu\neq 0$. Now $|x|$ again increases if x_0, x_1, x_2, x_3 are replaced by new numbers x_0', x_1', x_2', x_3', where

$$x_0'=x_0+\epsilon,\quad x_\nu'=x_\nu-\epsilon \quad \text{for} \quad x_0<\tfrac{1}{2},$$

and

$$x_1'=x_1+\epsilon,\quad x_\nu'=x_\nu-\epsilon \quad \text{for} \quad x_0=\tfrac{1}{2},$$

while the two other x's are left unchanged.

Also, if (32) is false and therefore $x_1>0$, then we may add ϵ to x_2 or x_3 and so increase $|x|$.

Finally, if (33) does not hold, then we can increase $|x|$ by replacing x_0 by $x_0+\epsilon$ and x_1 by $x_1-\epsilon$.

We now apply Theorem 1 to the set $\mathrm{d}(\Lambda)$. According to this theorem, the maximum of $|x|$ can be assumed only on those points of the set in which at least four of the boundary hyperplanes

$$L_1=x_3=0,\quad L_2=x_2-x_3=0,\quad L_3=x_1-x_2=0,\quad L_4=x_0-x_1=0,$$

$$L_5=x_0-\tfrac{1}{2}=0,\quad L_6=x_0+x_1+x_2+x_3-1=0,\quad L_7=x_0+x_1-\Lambda=0$$

intersect. The results (31)–(33) exclude all possibilities except the following ten cases:

$$L_1 = L_2 = L_3 = L_7 = 0, \quad x = \Lambda, \quad |x|^2 = \Lambda^2, \qquad \text{valid for} \quad 0 \leqslant \Lambda \leqslant \tfrac{1}{2},$$

$$L_2 = L_3 = L_4 = L_7 = 0, \quad x = \tfrac{1}{2}\Lambda(1+i_1+i_2+i_3), \quad |x|^2 = \Lambda^2,$$

$$\text{valid for} \quad 0 \leqslant \Lambda \leqslant \tfrac{1}{2},$$

$$L_2 = L_3 = L_5 = L_7 = 0, \quad x = \tfrac{1}{2} + (\Lambda - \tfrac{1}{2})(i_1+i_2+i_3), \quad |x|^2 = 3\Lambda^2 - 3\Lambda + 1,$$

$$\text{valid for} \quad \tfrac{1}{2} \leqslant \Lambda \leqslant \tfrac{2}{3},$$

$$L_1 = L_3 = L_6 = L_7 = 0, \quad x = (2\Lambda - 1) + (1-\Lambda)(i_1+i_2), \quad |x|^2 = 6\Lambda^2 - 8\Lambda + 3,$$

$$\text{valid for} \quad \tfrac{2}{3} \leqslant \Lambda \leqslant \tfrac{3}{4},$$

$$L_1 = L_5 = L_6 = L_7 = 0, \quad x = \tfrac{1}{2} + (\Lambda - \tfrac{1}{2})i_1 + (1-\Lambda)i_2, \quad |x|^2 = 2\Lambda^2 - 3\Lambda + \tfrac{3}{2},$$

$$\text{valid for} \quad \tfrac{3}{4} \leqslant \Lambda \leqslant 1,$$

$$L_2 = L_3 = L_6 = L_7 = 0, \quad x = \frac{(3\Lambda-1) + (1-\Lambda)(i_1+i_2+i_3)}{2},$$

$$|x|^2 = 3\Lambda^2 - 3\Lambda + 1, \quad \text{valid for} \quad \tfrac{1}{2} \leqslant \Lambda \leqslant$$

$$L_2 = L_5 = L_6 = L_7 = 0, \quad x = \frac{1 + (2\Lambda-1)i_1 + (1-\Lambda)(i_2+i_3)}{2},$$

$$|x|^2 = \tfrac{3}{2}\Lambda^2 - 2\Lambda + 1, \quad \text{valid for} \quad \tfrac{2}{3} \leqslant \Lambda \leqslant 1$$

$$L_3 = L_4 = L_6 = L_7 = 0, \quad x = \frac{\Lambda(1+i_1+i_2) + (2-3\Lambda)i_3}{2},$$

$$|x|^2 = 3\Lambda^2 - 3\Lambda + 1, \quad \text{valid for} \quad \tfrac{1}{2} \leqslant \Lambda \leqslant \tfrac{2}{3},$$

$$L_3 = L_5 = L_6 = L_7 = 0, \quad x = \frac{1 + (2\Lambda-1)(i_1+i_2) + (3-4\Lambda)i_3}{2},$$

$$|x|^2 = 6\Lambda^2 - 8\Lambda + 3, \quad \text{valid for} \quad \tfrac{2}{3} \leqslant \Lambda \leqslant \tfrac{3}{4},$$

$$L_4 = L_5 = L_6 = L_7 = 0, \quad x = \Sigma_0, \quad |x|^2 = \tfrac{1}{2}, \qquad \text{valid for} \quad \Lambda = 1.$$

The interval conditions for Λ must be satisfied if the points are to lie in the set $d(\Lambda)$. A trivial discussion shows that all 10 points satisfy the inequality (30), which is therefore true for all points of $d(\Lambda)$.

9. *An extremum problem.*

Let $\mathfrak{r}$ be a number in the interval

$$0 \leqslant \mathfrak{r} \leqslant \sqrt{\tfrac{1}{2}}, \tag{34}$$

and x a point in Δ of absolute value $\mathfrak{x}$. Put

$$m(x) = \min_{g,\,h} (|gx+\Sigma_0-h|), \tag{35}$$

where the minimum is extended over the 6 units

$$g = 1,\ i_1,\ \frac{1+i_1 \mp i_2 \mp i_3}{2}$$

and over all integral quaternions h. Obviously, $m(x)$ is a bounded continuous function of x; it therefore assumes a maximum value

$$M(\mathfrak{x}) = \max_{\substack{|x|=\mathfrak{x} \\ x \text{ in } \Delta}} m(x). \tag{36}$$

THEOREM 9. *The function $M(\mathfrak{x})$ satisfies the inequality*

$$M(\mathfrak{x}) \leqslant \sqrt{(\tfrac{1}{2}-\mathfrak{x}^2)}. \tag{37}$$

Proof. Since the g's are units,

$$m(x) = \min_{g,\,h} \left(|x-g^{-1}(h-\Sigma_0)|\right).$$

Now if g and h take all admissible values, then $g^{-1}(h-\Sigma_0)$, as is easily verified, runs over all SV Σ. Hence

$$m(x) = \min_{\Sigma} (|x-\Sigma|),$$

where the minimum sign extends over all SV. Let ϵ be one of the 24 units of K. Then with Σ also $\epsilon^{\mp 1}\Sigma$ runs over all SV; therefore

$$m(\epsilon x) = \min_{\Sigma} (|\epsilon x-\Sigma|) = \min_{\Sigma} (|x-\epsilon^{-1}\Sigma|) = \min_{\Sigma} (|x-\Sigma|) = m(x).$$

Hence, by Theorem 3, Part (b), we may restrict x to the set Δ^*, so that

$$M(\mathfrak{x}) = \max_{\substack{|x|=\mathfrak{x} \\ x \text{ in } \Delta^*}} \left\{\min_{\Sigma} (|x-\Sigma|)\right\}.$$

But then, by Theorem 4, Part (b), Σ_0 is the SV nearest to x, so that

$$M(\mathfrak{x}) = \max_{\substack{|x|=\mathfrak{x} \\ x \text{ in } \Delta^*}} (|x-\Sigma_0|).$$

Finally, $|x-\Sigma_0|$ remains unchanged if x_0 and x_1, or x_2 and x_3 are interchanged, or if we change the signs of x_2 or x_3 or both. Therefore

$$M(\mathfrak{x}) = \max_{T} (|x-\Sigma_0|),$$

the maximum being taken for the set T of all points x satisfying

$$|x| = \mathfrak{r}, \quad 0 \leqslant x_3 \leqslant x_2 \leqslant x_1 \leqslant x_0 \leqslant \tfrac{1}{2}, \quad x_0+x_1+x_2+x_3 \leqslant 1.$$

From this formula we get

$$M(\mathfrak{r})^2 = \max_T (\mathfrak{r}^2 - x_0 - x_1 + \tfrac{1}{2}) = \mathfrak{r}^2 + \tfrac{1}{2} - \min_T (x_0+x_1).$$

The statement follows therefore at once, since, by Theorem 8,

$$x_0+x_1 \geqslant 2\mathfrak{r}^2$$

for all points x such that

$$|x| \geqslant \mathfrak{r}, \quad 0 \leqslant x_3 \leqslant x_2 \leqslant x_1 \leqslant x_0 \leqslant \tfrac{1}{2}, \quad x_0+x_1+x_2+x_3 \leqslant 1.$$

This completes the proof.

It is possible to determine $M(\mathfrak{r})$ explicitly, namely:

$$M(\mathfrak{r})^2 = \begin{cases} \mathfrak{r}^2 + \frac{1}{2} - \mathfrak{r} & \text{for } 0 \leqslant \mathfrak{r} \leqslant \frac{1}{2}, \\ \mathfrak{r}^2 + \frac{1}{2} - \dfrac{3+\sqrt{(12\mathfrak{r}^2-3)}}{6} & \text{for } \frac{1}{2} \leqslant \mathfrak{r} \leqslant \sqrt{\frac{1}{3}}, \\ \mathfrak{r}^2 + \frac{1}{2} - \dfrac{4+\sqrt{(6\mathfrak{r}^2-2)}}{6} & \text{for } \sqrt{\frac{1}{3}} \leqslant \mathfrak{r} \leqslant \sqrt{\frac{3}{8}}, \\ \mathfrak{r}^2 + \frac{1}{2} - \dfrac{3+\sqrt{(8\mathfrak{r}^2-3)}}{4} & \text{for } \sqrt{\frac{3}{8}} \leqslant \mathfrak{r} \leqslant \sqrt{\frac{1}{2}}. \end{cases}$$

I omit the proof, since it requires no new ideas and since we do not make use of the result.

10. *A property of special Hermitian forms.*

We can now prove the following result:

THEOREM 10. *Let a be a number in the interval*

$$1 \leqslant a \leqslant \sqrt{2}, \tag{38}$$

let λ be the number

$$\lambda = \sqrt{\left(1 - \frac{1}{a^2}\right)}, \tag{39}$$

and let ξ be a quaternion in Δ of absolute value $|\xi| = \lambda$, and $H(x, y|\xi)$ the Hermitian form

$$H(x, y) = H(x, y|\xi) = aN(x - y\xi) + \frac{1}{a} N(y). \tag{40}$$

Then to any two quaternions x_0, y_0 there are two other quaternions x_1 and y_1, such that

$$x_1 \equiv x_0, \quad y_1 \equiv y_0, \quad H(x_1, y_1) \leqslant 1; \tag{41}$$

here the sign "$\leqslant$" may be replaced by "$<$", except in the case

$$x_0 = \Sigma_1, \quad y_0 = \Sigma_2, \quad a = 1, \quad \xi = 0, \quad H(x, y) = N(x) + N(y),$$

where Σ_1 and Σ_2 are two SV.

Proof. For every unit ϵ in K, the congruence $u \equiv v$ implies

$$\epsilon u \equiv \epsilon v \quad \text{and} \quad u\epsilon \equiv v\epsilon.$$

It has been proved in Theorems 2 and 3 that to every quaternion x there is a congruent one in Δ, and to every x in Δ there exists a unit ϵ such that ϵx lies in Δ^*. It is further true that to this x in Δ there is also a unit ϵ' such that $x\epsilon'$ belongs to Δ^*. For with x also the conjugate quaternion $\bar{x}$ lies in Δ. Therefore there is a unit $\bar{\epsilon}^*$ such that $\bar{\epsilon}^* \bar{x}$ belongs to Δ^*; it is easily verified that also

$$(\overline{\bar{\epsilon}^* \bar{x}})\, i_1 = x\epsilon^* i_1 = x\epsilon', \quad \text{where} \quad \epsilon' = \epsilon^* i_1,$$

is an element of Δ^*, and here ϵ' is a unit.

$H(x, y)$ satisfies the identities

$$H(\epsilon x, \epsilon y \mid \xi) = H(x, y \mid \xi),$$
$$H(x\epsilon, y \mid \xi\epsilon) = H(x, y \mid \xi).$$

From this and the preceding we may assume without loss of generality that

$$y_0 \quad \text{and} \quad z_0 = x_0 - y_0 \xi \quad \text{lie in} \quad \Delta^*.$$

Then by Theorem 2,

$$|y_0|^2 \leqslant \tfrac{1}{2} \quad \text{and} \quad |z_0|^2 \leqslant \tfrac{1}{2}.$$

Hence (41) is true in the stronger form with the sign "$<$" instead of "$\leqslant$" for $x_1 = x_0$ and $y_1 = y_0$, if

$$\text{either} \quad |y_0|^2 < \frac{2a - a^2}{2} \quad \text{or} \quad |z_0|^2 < \frac{2a-1}{2a^2}.$$

We exclude these two trivial cases and suppose that

$$|y_0|^2 \geqslant \frac{2a - a^2}{2}, \quad |z_0|^2 \geqslant \frac{2a-1}{2a^2}. \tag{42}$$

By § 6, the first inequality implies that y_0 belongs to $\delta(\tau_1)$, where τ_1 is determined by

$$\rho_1^2 = \frac{2a-a^2}{2}; \quad \text{hence} \quad \tau_1 = \frac{1-\sqrt{(8a-4a^2-3)}}{4}.$$

Therefore by Theorem 7, if g is one of the 6 quaternions

$$g = 1, \ i_1, \ \frac{1+i_1 \mp i_2 \mp i_3}{2},$$

then all 6 points

$$y_1^* = y_0 - g$$

satisfy the inequality

$$|y_1^*|^2 \leqslant \frac{1+2a-a^2-\sqrt{(8a-4a^2-3)}}{2}.$$

By the change of y_0 into y_1^*, z_0 is transformed into the 6 numbers

$$z_1^* = z_0 + g\xi = (z_0 - \Sigma_0) + (\Sigma_0 + g\xi).$$

Let g and an integral quaternion h be chosen such that

$$|g\xi + \Sigma_0 - h|$$

is a minimum and therefore equal to $m(\xi)$ as defined in (35); then put

$$x_1 = x_0 - h, \quad y_1 = y_0 - g, \quad z_1 = x_1 - y_1\xi = z_0 + g\xi - h.$$

Hence, by the inequality for y_1^*,

$$\frac{1}{a} N(y_1) = \frac{1+2a-a^2-\sqrt{(8a-4a^2-3)}}{2a} = A, \text{ say.} \tag{43}$$

As was proved in Theorem 9,

$$m(\xi) = |g\xi + \Sigma_0 - h| \leqslant M(\lambda) \leqslant \sqrt{(\tfrac{1}{2} - \lambda^2)} = \sqrt{\left(\frac{1}{a^2} - \frac{1}{2}\right)}.$$

By the second inequality (42), z_0 belongs to the set $\delta(\tau_2)$, where τ_2 is defined by

$$\rho_2^2 = \frac{2a-1}{2a^2}; \quad \text{hence} \quad \tau_2 = \frac{a-\sqrt{(8a-4-3a^2)}}{4a}.$$

Now the 16 points (20) in which sets of 4 of the boundary hyperplanes of $\delta(\tau)$ intersect, all have the same distance $\tau\sqrt{2}$ from the centre Σ_0 of this set. Hence, by Theorem 1, this is the maximum distance of a point of $\delta(\tau)$ from Σ_0. Therefore, in particular,

$$|z_0 - \Sigma_0| \leqslant \tau_2 \sqrt{2} = \frac{a-\sqrt{(8a-4-3a^2)}}{4a} \sqrt{2}.$$

Hence

$$|z_1| \leqslant |g\xi+\Sigma_0-h|+|z_0-\Sigma_0| \leqslant \sqrt{\left(\frac{1}{a^2}-\frac{1}{2}\right)}+\frac{a-\sqrt{(8a-4-3a^2)}}{4a}\sqrt{2},$$

so that

$$aN(x_1-y_1\xi) \leqslant a\left(\frac{1}{2a}\sqrt{(4-2a^2)}+\frac{a-\sqrt{(8a-4-3a^2)}}{4a}\sqrt{2}\right)^2 = B, \quad \text{say.} \tag{44}$$

With the values (43) and (44) for A and B, we have proved that

$$H(x_1, y_1) \leqslant A+B. \tag{45}$$

If $a=1$ and therefore $\xi=0$, then $A=B=\frac{1}{2}$, so that (41) is satisfied. By Theorem 2, the sign of equality holds only if $N(x_1)=N(y_1)=\frac{1}{2}$, *i.e.* if both x_0 and y_0 are SV.

Now let $1<a\leqslant\sqrt{2}$. Then we show that A and B are both less than $\frac{1}{2}$. In the case of A, we have

$$A = \tfrac{1}{2}+\frac{1+a-a^2-\sqrt{(8a-4a^2-3)}}{2a}$$

$$= \tfrac{1}{2}+\frac{(a-1)(a^3-a^2+2a-4)}{2a\{1+a-a^2+\sqrt{(8a-4a^2-3)}\}}.$$

Here

$$a^3-a^2+2a-4 = -2+3(a-1)+2(a-1)^2+(a-1)^3$$

is an increasing function of a; therefore its maximum is assumed for $a=\sqrt{2}$, and is equal to

$$4\sqrt{2}-6<0.$$

This proves that $A<\frac{1}{2}$.

The inequality $B<\frac{1}{2}$ is equivalent to

$$\sqrt{(2B)}=\beta=\frac{1}{\sqrt{a}}\sqrt{(2-a^2)}+\frac{a-\sqrt{(8a-4-3a^2)}}{2\sqrt{a}}<1.$$

Now

$$\sqrt{(2-a^2)}-1=\frac{1-a^2}{\sqrt{(2-a^2)}+1} \leqslant 1-a^2=-(a+1)(a-1) \leqslant -2(a-1).$$

Hence

$$\frac{1}{\sqrt{a}}\sqrt{(2-a^2)}<1-\frac{2(a-1)}{\sqrt{a}}.$$

Also

$$\frac{a-\sqrt{(8a-4-3a^2)}}{2\sqrt{a}}=\frac{4(a-1)^2}{2\sqrt{a}\{a+\sqrt{(8a-4-3a^2)}\}} \leqslant \frac{2(a-1)^2}{a\sqrt{a}}.$$

Therefore finally

$$\beta < \left(1-\frac{2(a-1)}{\sqrt{a}}\right)+\frac{2(a-1)^2}{a\sqrt{a}} = 1-\frac{2(a-1)}{a\sqrt{a}} < 1,$$

which concludes the proof.

COROLLARY TO THEOREM 10. *The assertion* (41) *remains true, if* $H(x, y)$ *is replaced by the Hermitian form*

$$\bar{H}(x, y\,|\,\xi) = aN(x-\xi y)+\frac{1}{a}N(y).$$

For evidently $$\bar{H}(x, y\,|\,\xi) = H(\bar{x}, \bar{y}\,|\,\bar{\xi}).$$

CHAPTER II. HERMITIAN FORMS.

11. *Linear transformations.*

The quaternion field K is associative; therefore the composition of linear transformations

$x = a_{11}x'+a_{12}y'$, $y = a_{21}x'+a_{22}y'$, or in matrix form $\begin{pmatrix} x \\ y \end{pmatrix} = \Omega \begin{pmatrix} x' \\ y' \end{pmatrix}$, where $\Omega = \begin{pmatrix} a_{11} a_{12} \\ a_{21} a_{22} \end{pmatrix}$,

satisfies the ordinary rules. The determinant of commutative algebra, however, loses its meaning, and has to be replaced by the expression†

$$(46)\quad d(\Omega) = a_{11}\bar{a}_{11}a_{22}\bar{a}_{22}+a_{12}\bar{a}_{12}a_{21}\bar{a}_{21}-a_{11}\bar{a}_{21}a_{22}\bar{a}_{12}-a_{12}\bar{a}_{22}a_{21}\bar{a}_{11},$$

which we call the *determinant of* Ω. As we shall see, $d(\Omega)$ *is either positive or zero*; in the first case Ω is called *regular*. For example, if

$$(47)\qquad S(\lambda) = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad U(\tau) = \begin{pmatrix} \tau & 0 \\ 0 & 1 \end{pmatrix},$$

where λ and $\tau \neq 0$ are quaternions, then

$$d\big(S(\lambda)\big) = 1, \quad d(T) = 1, \quad d\big(U(\tau)\big) = \tau\bar{\tau},$$

so that all three matrices are regular. These matrices have inverses, namely

$$(48)\qquad S(\lambda)^{-1} = S(-\lambda), \quad T^{-1} = T, \quad U(\tau)^{-1} = U(\tau^{-1}).$$

† For the general theory of generalized determinants in the quaternion field see the beautiful paper of E. Study, *Acta Mathematica*, 42 (1920), 1–61.

Every regular matrix can be written as a finite product of matrices S, T, and U. For, since $d(\Omega) \neq 0$, we have either $a_{12} \neq 0$, or $a_{12} = 0$ and $a_{11} \neq 0$, $a_{22} \neq 0$. In the first case

$$\Omega = S(\lambda_1)\, TU(\tau_1)\, TU(\tau_2)\, TS(\lambda_2),$$

where

$$\tau_1 = a_{21}, \quad \tau_2 = a_{12} - a_{11} a_{21}^{-1} a_{22}, \quad \lambda_1 = a_{11} a_{21}^{-1}, \quad \lambda_2 = a_{21}^{-1} a_{22};$$

and in the second case

$$\Omega = U(a_{11})\, S(a_{11}^{-1} a_{12} a_{22}^{-1})\, TU(a_{22})\, T.$$

From this factorization of Ω and from (48), we see that every regular matrix has an inverse, obtained by taking the inverses of the factors in the reversed order.

THEOREM 11. *If Ω_1 and Ω_2 are regular matrices, and $\Omega_3 = \Omega_1 \Omega_2$, then*

$$d(\Omega_3) = d(\Omega_1)\, d(\Omega_2). \tag{49}$$

Proof. It suffices to consider the cases in which

$$\Omega_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is an arbitrary regular matrix, and Ω_2 is one of the three matrices $S(\lambda)$, T or $U(\tau)$. The assertion is obvious in the two last cases. In the first case,

$$\Omega_3 = \Omega_1 S(\lambda) = \begin{pmatrix} a_{11}, & a_{11}\lambda + a_{12} \\ a_{21}, & a_{21}\lambda + a_{22} \end{pmatrix},$$

hence

$$d(\Omega_3) = a + b + c + d,$$

where $d = d(\Omega_1)$ and

$$a = a_{11}\bar{a}_{11}a_{21}\lambda\bar{\lambda}\bar{a}_{21} + a_{11}\lambda\bar{\lambda}\bar{a}_{11}a_{21}\bar{a}_{21} - a_{11}\bar{a}_{21}a_{21}\lambda\bar{\lambda}a_{11} - a_{11}\lambda\bar{\lambda}\bar{a}_{21}a_{21}\bar{a}_{11},$$

$$b = a_{11}\bar{a}_{11}a_{21}\lambda\bar{a}_{22} + a_{11}\lambda\bar{a}_{12}a_{21}\bar{a}_{21} - a_{11}\bar{a}_{21}a_{21}\lambda\bar{a}_{12} - a_{11}\lambda\bar{a}_{22}a_{21}\bar{a}_{11},$$

$$c = a_{11}\bar{a}_{11}a_{22}\bar{\lambda}\bar{a}_{21} + a_{12}\bar{\lambda}\bar{a}_{11}a_{21}\bar{a}_{21} - a_{11}\bar{a}_{21}a_{22}\bar{\lambda}a_{11} - a_{12}\bar{\lambda}\bar{a}_{21}a_{21}\bar{a}_{11}.$$

Since real factors are commutative, obviously $a = 0$. Further,

$$b + c = N(a_{11})\, S(a_{21}\lambda\bar{a}_{22}) + N(a_{21})\, S(a_{11}\lambda\bar{a}_{12}) - N(a_{21})\, S(a_{11}\lambda\bar{a}_{12}) - S(a_{11}\lambda\bar{a}_{22}a_{21}\bar{a}_{11}).$$

This expression vanishes, since $S(ab) = S(ba)$, and therefore the last term†

$$S(a_{11}\lambda\bar{a}_{22}\,.\,a_{21}\bar{a}_{11}) = S(a_{21}\bar{a}_{11}\,.\,a_{11}\lambda\bar{a}_{22}) = N(a_{11})\,S(a_{21}\lambda\bar{a}_{22}).$$

From Theorem 11, the inequality $d(\Omega) > 0$ follows at once, since it holds for the matrices S, T, U. We also see that the product of regular matrices is again regular; they therefore form a group.

12. *The modular group.*

Let Γ be the set of all matrices Ω of determinant 1 and with elements which are integral quaternions. By the last theorem, the product of two elements of Γ belongs again to this set. The set contains the unit element $S(0) = E$. As we now prove, it contains with every matrix Ω also the inverse matrix Ω^{-1}, and therefore it is a group, the *modular group in* K.

In order to establish the existence of Ω^{-1}, it suffices to show that Ω can be written as a product of a finite number of factors $S(\lambda)$, T, and $U(\tau)$, where λ is an integral quaternion, and τ a unit. This can be done just as for the modular group in the rational field; all we need know is that the Euclidian algorithm holds in K‡.

If the element a_{21} in Ω vanishes, then, since $d(\Omega) = 1$, both a_{11} and a_{22} are units; hence

$$\Omega = U(a_{11})\,S(a_{11}^{-1}a_{12}a_{22}^{-1})\,TU(a_{22})\,T$$

is a factorization of the required kind.

If, however, $a_{21} \neq 0$, then by Theorem 2 there is an integral quaternion λ such that $a_{21}^{-1}a_{22} - \lambda$ lies in Δ; therefore

$$|a_{21}^{-1}a_{22} - \lambda| \leqslant \sqrt{\tfrac{1}{2}}, \quad i.e., \quad N(a_{22} - a_{21}\lambda) \leqslant \tfrac{1}{2}N(a_{21}).$$

We now have

$$\Omega = \Omega_1 TS(\lambda), \qquad \Omega_1 = \begin{pmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{pmatrix}, \quad \text{where} \quad N(a'_{21}) \leqslant \tfrac{1}{2}N(a_{21}),$$

since obviously $$a'_{21} = a_{22} - a_{21}\lambda.$$

If still $a'_{21} \neq 0$, then there is an integral quaternion λ' such that

$$\Omega_1 = \Omega_2 TS(\lambda'), \qquad \Omega_2 = \begin{pmatrix} a''_{11} & a''_{12} \\ a''_{21} & a''_{22} \end{pmatrix}, \quad \text{where} \quad N(a''_{21}) \leqslant \tfrac{1}{2}N(a'_{21}).$$

† We note that $S(\alpha\beta) = S(\beta\alpha) = \alpha S(\beta)$, if α is real.

‡ *Loc. cit.*, footnote †, p. 435.

Repeating this process, if necessary, we finally come to a matrix in which the coefficient with indices (21) vanishes, thus to the case already dealt with.

The factorization of Ω allows us to write down the inverse merely by taking the inverses of the factors in the reversed order.

13. *The discriminant of a Hermitian form.*

An expression

$$H(x, y) = \bar{x}Ax+\bar{x}By+\bar{y}\bar{B}x+\bar{y}Cy, \tag{50}$$

where A and C are real numbers and B and $\bar{B}$ are conjugate quaternions, is called a *Hermitian form in* K; we call

$$D = AC-B\bar{B} \tag{51}$$

its *discriminant*. Evidently, $H(x, y)$ can be written as

$$H(x, y) = AN(x-\xi y)+\frac{D}{A}N(y), \quad \text{where} \quad \xi = -\frac{B}{A}. \tag{52}$$

We consider only *positive definite forms* $H(x, y)$, *i.e.* we assume that $H(x, y) > 0$ for $N(x)+N(y) > 0$. The last formula shows that this is the case if, and only if,

$$A > 0 \quad \textit{and} \quad D > 0.$$

A linear transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \Omega \begin{pmatrix} x \\ y \end{pmatrix} \tag{53}$$

changes $H(x, y)$ into a new positive definite Hermitian form

$$H'(x, y) = \bar{x}A'x+\bar{x}B'y+\bar{y}\bar{B}'x+\bar{y}C'y$$

of discriminant

$$D' = A'C'-B'\bar{B}'.$$

Here

$$D' = D\,d(\Omega). \tag{54}$$

For by the matrix factorization in § 11, it suffices to prove this equation for the three matrices $S(\lambda)$, T, $U(\tau)$, and it is easily verified in these three cases.

If the matrix Ω in (53) is an element of Γ, then $H'(x, y)$ is called *equivalent to* $H(x, y)$, in signs $H' \sim H$. Since Γ is a group, this relation

has the properties

$$H \sim H;$$

$$\text{if } H' \sim H, \text{ then } H \sim H';$$

$$\text{if } H' \sim H \text{ and } H'' \sim H', \text{ then } H'' \sim H.$$

By (54), *equivalent forms have equal discriminants.*

Equivalent forms evidently represent the same numbers when the arguments run over all integral quaternions. It is further clear from the definition of equivalence that, if $H' \sim H$, then both coefficients A' and C' are of the form $H(x, y)$ with integral x and y. Hence the two sets consisting respectively of the first coefficients A', and of the last coefficients C', of all forms $H' \sim H$, contain only *a finite number* of elements less than a given constant.

14. *The reduction of Hermitian forms.*

The form $H(x, y)$ is called *reduced* if

$$H(x, y) \geqslant \begin{cases} A \text{ for integral } x \text{ and } y \text{ with } N(x)+N(y) > 0, \\ C \text{ for integral } x \text{ and } y \text{ with } N(y) > 0. \end{cases} \tag{55}$$

THEOREM 12. *The form $H(x, y)$ is reduced if, and only if,*

$$A \leqslant C, \text{ and } \xi = -\frac{B}{A} \text{ is an element of } \Delta. \tag{56}$$

Proof. If (55) holds, then $C = H(0, 1) \geqslant A$; further, by (52), for all integral x,

$$0 \leqslant H(x, 1) - C = AN(x-\xi) - \frac{B\bar{B}}{A} = A\Big(N(x-\xi) - N(\xi)\Big),$$

so that ξ lies in Δ.

If (56) is true, then

$$H(x, 0) = N(x)\, H(1, 0) \geqslant A$$

for integral $x \neq 0$. Further, for every unit ϵ,

$$H(x, \epsilon) - C = H(x\epsilon^{-1}, 1) - C = A\Big(N(x\epsilon^{-1} - \xi) - N(\xi)\Big) \geqslant 0,$$

since ξ lies in Δ; and, for $N(y) \geqslant 2$,

$$\begin{aligned} H(x, y) - C &= AN(x-\xi y) + C\Big(N(y)-1\Big) - \frac{B\bar{B}}{A}\, N(y) \\ &\geqslant 0 + \frac{C}{2}\, N(y) - AN(\xi)\, N(y) \geqslant AN(y)\Big(\tfrac{1}{2} - N(\xi)\Big) \geqslant 0, \end{aligned}$$

since $A \leqslant C$ and $N(\xi) \leqslant \frac{1}{2}$.

THEOREM 13. *If $H(x, y)$ is not reduced, then there is an equivalent form $H^*(x, y)$ such that*

$$\text{either} \quad A^* < A, \quad \text{or} \quad A^* = A, \quad A^* \leqslant C^* < C.$$

Proof. If $A > C$, then we may put $H^*(x, y) = H(y, x)$, so that $A^* = C < A$. Suppose therefore that $A \leqslant C$, but that $\xi = -\frac{B}{A}$ does not lie in Δ. Then let λ be an integral quaternion such that $\xi - \lambda$ lies in Δ. The transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} \to S(\lambda) \begin{pmatrix} x \\ y \end{pmatrix}$$

changes $H(x, y)$ into a form $H'(x, y)$ in which

$$A' = A, \quad B' = A\lambda + B, \quad C' = H(\lambda, 1).$$

If now $C' < A$, then we may put $H^*(x, y) = H'(y, x)$, so that

$$A^* = C' < A' = A.$$

If, however, $C' \geqslant A$, then put $H^*(x, y) = H'(x, y)$, so that

$$C^* - C = C' - C = A\big(N(\xi - \lambda) - N(\xi)\big) < 0,$$

since $\xi - \lambda$ but not ξ lies in Δ.

THEOREM 13. *To every form $H(x, y)$ there exists an equivalent form $H^*(x, y)$ which is reduced.*

Proof. If $H(x, y)$ is not itself reduced, then by the last theorem we can find a sequence of equivalent forms

$$H_1(x, y), \; H_2(x, y), \; \ldots$$

where

$$\text{either} \quad A_1 < A, \quad \text{or} \quad A_1 = A, \quad A_1 \leqslant C_1 < C,$$

$$\text{either} \quad A_2 < A_1, \quad \text{or} \quad A_2 = A_1, \quad A_2 \leqslant C_2 < C_1,$$

etc. Since by the last paragraph there are only a finite number of possible values of the first and the last coefficient of a form equivalent to $H(x, y)$ not exceeding a given value, this sequence of forms necessarily terminates after a finite number of terms; its last element is the wanted reduced form $H^*(x, y)$.

15. *The minima of a Hermitian form.*

Let $$\mathrm{X} = \begin{pmatrix} x_1 x_2 \\ y_1 y_2 \end{pmatrix}$$

be a matrix of determinant different from zero, and with elements which are integral quaternions; further let $H(x, y)$ be a positive definite Hermitian form, and $H^*(x, y)$ a reduced form equivalent to $H(x, y)$. The two expressions

$$(57) \qquad \begin{cases} M_1 = \min\limits_{\mathrm{X}} \left\{ \min \left(H(x_1, y_1),\ H(x_2, y_2) \right) \right\}, \\ M_2 = \min\limits_{\mathrm{X}} \left\{ \max \left(H(x_1, y_1),\ H(x_2, y_2) \right) \right\}, \end{cases}$$

where X takes all admissible values, are called the *first and second minimum of* $H(x, y)$. It is clear that M_1 is also the minimum of $H(x, y)$ when x and y run over all pairs of integral quaternions except the pair $x = y = 0$.

THEOREM 14. *If A^* and C^* are the first and the last coefficient of $H^*(x, y)$, then the two minima of $H(x, y)$ are given by*

$$(58) \qquad M_1 = A^*, \quad M_2 = C^*.$$

Proof. Let $$\begin{pmatrix} x \\ y \end{pmatrix} = \Omega \begin{pmatrix} x^* \\ y^* \end{pmatrix}$$

be the transformation in Γ for which

$$H(x, y) = H^*(x^*, y^*),$$

and let $$\begin{pmatrix} x^* \\ y^* \end{pmatrix} = \Omega^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$$

be its inverse. Put

$$\mathrm{X}^* = \Omega^{-1}\mathrm{X} = \begin{pmatrix} x_1^* x_2^* \\ y_1^* y_2^* \end{pmatrix}, \quad \text{so that} \quad \mathrm{X} = \Omega \mathrm{X}^*.$$

The elements of X^* are also integral quaternions, and the determinant $d(\mathrm{X}^*) = d(\mathrm{X}) \neq 0$. Hence the elements in one row or column of X^* are not both zero. Evidently,

$$H(x_1, y_1) = H^*(x_1^*, y_1^*), \quad H(x_2, y_2) = H^*(x_2^*, y_2^*).$$

Hence, by the definition of reduction,

$$\min \left(H(x_1, y_1), H(x_2, y_2) \right) \geqslant A^*, \quad \max \left(H(x_1, y_1), H(x_2, y_2) \right) \geqslant C^*.$$

If, in particular,

$$X^* = E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \Omega,$$

then

$$H(x_1, y_1) = H^*(1, 0) = A^*, \quad H(x_2, y_2) = H^*(0, 1) = C^*,$$

so that the last formulae hold with the sign of equality.

THEOREM 15. *The minima M_1 and M_2 of $H(x, y)$ satisfy the inequalities*

$$M_1 \leqslant \sqrt{(2D)}, \quad D \leqslant M_1 M_2 \leqslant 2D. \tag{59}$$

In particular, $M_1 = \sqrt{(2D)}$, and also $M_1 M_2 = 2D$, if, and only if, $A^* = C^*$ and ξ is one of the 24 SV of Δ. Further, $M_1 M_2 = D$ if, and only if, $B^* = 0$.

Proof. We have

$$D = A^* C^* - B^* \bar{B}^* \leqslant A^* C^*,$$

with equality if, and only if, $B^* = 0$. Further by Theorem 12, $A^* \leqslant C^*$ and $\xi = -\dfrac{B^*}{A^*}$ lies in Δ, so that $N(\xi) \leqslant \frac{1}{2}$ and therefore

$$B^* \bar{B}^* \leqslant \tfrac{1}{2} A^{*2} \leqslant \tfrac{1}{2} A^* C^*,$$

$$D = A^* C^* - B^* \bar{B}^* \geqslant A^* C^* - \tfrac{1}{2} A^* C^* = \tfrac{1}{2} A^* C^*.$$

If the sign of equality is to hold in the last formula, then $A^* = C^*$ and $N(\xi) = \frac{1}{2}$. The assertion follows therefore at once from Theorem 2 and Theorem 14.

Since Δ is invariant for the 24 transformations (6), all forms $H(x, y)$ satisfying $M_1 = \sqrt{(2D)}$ are equivalent to

$$H_0(x, y) = \sqrt{(2D)}\,(\bar{x}x - \bar{x}\,\Sigma_0 y - \bar{y}\,\bar{\Sigma}_0 x + \bar{y}y).$$

This form assumes its minimum value $M_1 = M_2 = \sqrt{(2D)}$ for $9 \times 24 = 216$ different pairs of integral x, y. For let ϵ be one of the 24 units of K; then the minimum is assumed for $(x, y) = (\epsilon, 0)$, and also for $(x, y) = (x, \epsilon)$ where now x is one of the 8 integral quaternions for which

$$N(x - \Sigma_0 \epsilon) = \tfrac{1}{2}.$$

The part of the last theorem referring to M_1 is due to A. Speiser†, who used an entirely different method in his proof.

† *Journal für Math.*, 167 (1932), 88–97, in particular p. 97.

16. *The geometrical representation of Hermitian forms.*

Let $H(x, y)$ be a Hermitian form of discriminant $D = 1$; put

$$\xi = -\frac{B}{A}, \quad \eta = \frac{1}{A}, \tag{60}$$

so that

$$A = \frac{1}{\eta}, \quad B = -\frac{\xi}{\eta}, \quad \bar{B} = -\frac{\bar{\xi}}{\eta}, \quad C = \frac{\xi\bar{\xi}+\eta^2}{\eta}. \tag{61}$$

We interpret the four real components of ξ together with the positive real number η as the rectangular coordinates of a point (ξ, η) in the part

$$R: \qquad \eta > 0$$

of five-dimensional Euclidian space, and then there is a one-to-one correspondence between the forms $H(x, y)$ and the points (ξ, η) of R.

If $H(x, y)$ of point (ξ, η) is equivalent to $H'(x', y')$ of point (ξ', η'), then these points are also called equivalent. Let

$$\begin{pmatrix} x \\ y \end{pmatrix} = \Omega \begin{pmatrix} x' \\ y' \end{pmatrix}, \quad \Omega = \begin{pmatrix} a_{11} a_{12} \\ a_{21} a_{22} \end{pmatrix}$$

be the transformation in Γ which changes the first form into the second. Then a simple calculation shows that

$$\begin{cases} \xi' = -\dfrac{(\bar{a}_{11}-\bar{a}_{21}\bar{\xi})(a_{12}-\xi a_{22})+\bar{a}_{21}a_{22}\eta^2}{(\bar{a}_{11}-\bar{a}_{21}\bar{\xi})(a_{11}-\xi a_{21})+\bar{a}_{21}a_{21}\eta^2}, \\ \eta' = \dfrac{\eta}{(\bar{a}_{11}-\bar{a}_{21}\bar{\xi})(a_{11}-\xi a_{21})+\bar{a}_{21}a_{21}\eta^2}, \end{cases} \tag{62}$$

with analogous formulae involving the elements of Ω^{-1} for the change-over from (ξ', η') to (ξ, η). The second equation shows that *R is transformed into itself by all elements of* Γ. We write (62) for shortness as $(\xi, \eta) = \Omega(\xi', \eta')$ or $(\xi', \eta') = \Omega^{-1}(\xi, \eta)$. In particular, for the generators of Γ,

$$\text{if} \quad \Omega = S(\lambda), \quad \text{then} \quad \xi' = \xi - \lambda, \quad \eta' = \eta,$$

$$\text{if} \quad \Omega = T, \quad \text{then} \quad \xi' = \frac{\bar{\xi}}{\xi\bar{\xi}+\eta^2}, \quad \eta' = \frac{\eta}{\xi\bar{\xi}+\eta^2},$$

$$\text{if} \quad \Omega = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix}, \quad \text{then} \quad \xi' = \epsilon_1^{-1}\xi\epsilon_2, \quad \eta' = \eta;$$

in the last formula, ϵ_1 and ϵ_2 are assumed to be units in K.

Now let F be the set of all points (ξ, η) which belong to reduced forms; by Theorems 2 and 12, F is determined by

$$(63) \qquad |\xi| \leqslant |\xi - g| \text{ for all integral quaternions } g; \quad \xi\bar{\xi} + \eta^2 \geqslant 1.$$

The point (ξ, η) is an *inner point* of F if

$$|\xi| < |\xi - g| \text{ for all integral quaternions } g \neq 0; \quad \xi\bar{\xi} + \eta^2 > 1;$$

denote by F_0 the set of all inner points of F. The proof of Theorem 12 shows that the form $H(x, y)$ belongs to a point of F_0 if, and only if, for all integral pairs (x, y),

$$(64) \qquad H(x, y) > \begin{cases} A & \text{when} \quad (x, y) \neq (0, 0), \quad \neq (\epsilon, 0), \\ C & \text{when} \quad y \neq 0 \quad \text{and} \quad (x, y) \neq (0, \epsilon), \end{cases}$$

where ϵ denotes the 24 units of K; in particular $A < C$.

Suppose now that $H(x, y)$ belongs to an inner point of F and that $H'(x', y')$ is a reduced form equivalent to $H(x, y)$. Let

$$\begin{pmatrix} x \\ y \end{pmatrix} = \Omega \begin{pmatrix} x' \\ y' \end{pmatrix}, \quad \Omega = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

be the transformation in Γ which changes one form into the other. Then by Theorem 14,

$$A = A', \quad C = C';$$

hence

$$A = H(a_{11}, a_{21}), \quad C = H(a_{12}, a_{22}).$$

By (64), the first formula requires that $a_{11} = \epsilon_1$, $a_{21} = 0$, the second one that $a_{12} = 0$, $a_{22} = \epsilon_2$, where ϵ_1 and ϵ_2 are any two units; therefore the transformation is of the type

$$(65) \qquad \Omega = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix}.$$

It follows that these are the only transformations which change at least one inner point of F into a point of F. Since the inverse of Ω has the same form, all 24×24 transformations (65) leave the set F_0 invariant.

Denote by $\Omega^{-1}F$ (and $\Omega^{-1}F_0$) the sets of all points (ξ', η') for which $(\xi, \eta) = \Omega(\xi', \eta')$ lies in F (in F_0). The last results show that F_0 and the transformed set $\Omega^{-1}F_0$ either are identical, or have no common points. Hence, by the group property of Γ, any two sets $\Omega_1^{-1}F_0$ and $\Omega_2^{-1}F_0$ either coincide or are without common points. Therefore, finally, two sets $\Omega_1^{-1}F$ and $\Omega_2^{-1}F$ either have no common inner points, or they

are identical. This shows that the sets $\Omega^{-1}F$ together fill up the space R without overlapping, and also without gaps since to every point there is an equivalent one in F.

If ϵ_1 and ϵ_2 are two units, and λ is an integral quaternion, then the set

$$\Omega^{-1}F, \quad \text{where} \quad \Omega = \begin{pmatrix} \epsilon_1 & \lambda \\ 0 & \epsilon_2 \end{pmatrix},$$

contains points with arbitrarily large η, since this is true for F, and the transformation leaves η invariant. When, however, $\Omega = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is an element of Γ with $a_{21} \neq 0$, then η is bounded for all points in $\Omega^{-1}F$. There are, however, points with arbitrarily small positive η in the set $\Omega^{-1}F$, as seen from (62), and these satisfy

$$\lim_{\eta \to 0} \xi = a_{21}^{-1} a_{22}. \tag{66}$$

It follows that if a continuous curve in R tends to a point $(\alpha, 0)$ in $\eta = 0$, where α is an irrational quaternion (*i.e.* not of the form $a^{-1}b$ with integral quaternions $a \neq 0$ and b), then this curve passes through an infinity of different sets $\Omega^{-1}F$. For if it passed only through a finite number, then it would ultimately lie in one set and so by (66) tend to a point $(a_{12}^{-1} a_{22}, 0)$.

17. *On special Hermitian forms.*

THEOREM 16. *Let* $\mathrm{A} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ *be a matrix of determinant* $d(\mathrm{A}) = 1$ *with elements in* K, t *a positive parameter,* $F_t(x, y)$ *the Hermitian form of discriminant* 1

$$F_t(x, y) = tN(\alpha x + \beta y) + \frac{1}{t} N(\gamma x + \delta y), \tag{67}$$

and

$$F_t^*(x, y) = \bar{x} A_t x + \bar{x} B_t y + \bar{y} \bar{B}_t x + \bar{y} C_t y \tag{68}$$

the reduced form equivalent to $F_t(x, y)$. *Then, for at least one value of* t,

$$A_t = C_t.$$

Further, if $\alpha^{-1}\beta$ *is not a rational quaternion, then there are arbitrarily large* t *with this property.*

Proof. Since

$$tF_t(x, y) = \bar{x}(\bar{\alpha}at^2 + \bar{\gamma}\gamma)x + \bar{x}(\bar{\alpha}\beta t^2 + \bar{\gamma}\delta)y + \bar{y}(\bar{\beta}at^2 + \bar{\delta}\gamma)x + \bar{y}(\bar{\beta}\beta t^2 + \bar{\delta}\delta)y,$$

the point (ξ, η) corresponding to $F_t(x, y)$ is given by

$$\xi = -\frac{\bar{\alpha}\beta t^2 + \bar{\gamma}\delta}{\bar{\alpha}at^2 + \bar{\gamma}\gamma}, \quad \eta = \frac{t}{\bar{\alpha}at^2 + \bar{\gamma}\gamma}, \quad \xi\bar{\xi} + \eta^2 = \frac{\bar{\beta}\beta t^2 + \bar{\delta}\delta}{\bar{\alpha}at^2 + \bar{\gamma}\gamma}.$$

If t assumes all positive values, then (ξ, η) describes the semicircle **C** which is perpendicular to the hyperplane $\eta = 0$ at the two points $(-\alpha^{-1}\beta, 0)$ and $(-\gamma^{-1}\delta, 0)$. For $t \to \infty$, (ξ, η) tends to the first point $(-\alpha^{-1}\beta, 0)$; hence, by the preceding paragraph, if $\alpha^{-1}\beta$ is irrational, then **C** passes through an infinity of different sets $\Omega^{-1}F$. (It is possible that just one of the two numbers α and γ vanishes; then **C** degenerates into a straight line perpendicular to $\eta = 0$.)

Suppose that a point $P: t = \tau$ of **C**, lies in the set $\Omega^{-1}F$. The transformation $\begin{pmatrix} x \\ y \end{pmatrix} \to \Omega \begin{pmatrix} x \\ y \end{pmatrix}$ changes $\Omega^{-1}F$ into F, **C** into a new semicircle or straight line **C*** perpendicular to $\eta = 0$, and the point P on **C** into a point $P^*(\xi^*, \eta^*)$ on **C***. Then P^* lies in F, so that

$$\xi^* \text{ belongs to } \Delta; \quad \xi^*\bar{\xi}^* + \eta^2 \geqslant 1.$$

Evidently P^* corresponds to the reduced form $F_\tau^*(x, y)$. The point divides **C*** into two parts; let c^* be that part of **C*** for whose points

$$\eta \leqslant \eta^*.$$

Let $g = g(\xi)$ be any integral quaternion such that $\xi - g$ lies in Δ. If there is more than one quaternion g with this property, then, by the definition of Δ, $N(\xi - g)$ has the same value for all of them.

Denote now by S the hypersurface in R defined by

$$(\xi - g)(\bar{\xi} - \bar{g}) + \eta^2 = 1. \tag{69}$$

Then η is a continuous single-valued function of ξ, and so S is a continuous and connected hypersurface; it is clearly transformed into itself by all translations of R,

$$\begin{pmatrix} x \\ y \end{pmatrix} \to S(\lambda) \begin{pmatrix} x \\ y \end{pmatrix}, \quad i.e. \quad \xi \to \xi + \lambda, \quad \eta \to \eta, \tag{70}$$

where λ runs over all integral quaternions.

The point P^* belongs to F and therefore either lies on S or has a larger coordinate η^* than the point of S with the same value ξ^* of ξ. In the second case the arc c^* of $\mathbf{C}^*$ intersects S in at least one point, since it joins P^* to a point in $\eta = 0$. By a suitable translation (70), this point of c^* changes into one on S and inside F. It follows that there is a transformation in Γ which changes $\mathbf{C}$ into a semicircle or straight line intersecting S in a point of F. If $t = t_0$ is the parameter value belonging to this point of intersection, then, for the reduced form (68), $A_{t_0} = C_{t_0}$, as follows from (61) since $\xi - g$ in (69) has now been replaced by ξ.

If $\alpha^{-1}\beta$ is irrational, then there are an infinity of ways of changing $\mathbf{C}$ into a semicircle or straight line which intersects S in a point of F. For now $\mathbf{C}$ passes through an infinity of different sets $\Omega^{-1}F$ in the neighbourhood of $(\alpha^{-1}\beta, 0)$. A transformation which changes one of these sets into F transforms at most a finite number of the others into sets $S(\lambda)^{-1}F$. This proves the assertion.

18. *The product of two linear polynomials.*

THEOREM 17. *Let* $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ *be a matrix of determinant* $d(A) = 1$ *with elements in* K. *Corresponding to any quaternions* x_0 *and* y_0 *there are two other quaternions,* x_1 *and* y_1, *such that*

$$\text{(A)} \qquad x_1 \equiv x_0, \quad y_1 \equiv y_0, \quad |(\alpha x_1 + \beta y_1)(\gamma x_1 + \delta y_1)| \leqslant \tfrac{1}{2}.$$

If $\alpha^{-1}\beta$ *is irrational, then* (A) *has a solution with arbitrarily small value of* $|\alpha x_1 + \beta y_1|$.

Proof. Consider the forms (67) and (68). By Theorem 16, there is a value $t = t_0$ for which $A_{t_0} = C_{t_0}$. If $\alpha^{-1}\beta$ is irrational, then t_0 may be taken arbitrarily large. *Let*

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \Omega \begin{pmatrix} x \\ y \end{pmatrix}, \quad \Omega = \begin{pmatrix} a_{11} a_{12} \\ a_{21} a_{22} \end{pmatrix}$$

be the transformation in Γ *which changes* $F_{t_0}(x, y)$ *into* $F^*_{t_0}(x', y')$, *and let*

$$x_0' = a_{11}x_0 + a_{12}y_0, \quad y_0' = a_{21}x_0 + a_{22}y_0.$$

By the corollary to Theorem 10 and by (52), there are two quaternions x_1' and y_1' such that

$$x_1' \equiv x_0', \quad y_1' \equiv y_0', \quad F^*_{t_0}(x_1', y_1') \leqslant 1.$$

Then, if x_1 and y_1 are the quaternions defined by

$$x_1' = a_{11}x_1 + a_{12}y_1, \quad y_1' = a_{21}x_1 + a_{22}y_1,$$

obviously

$$x_1 \equiv x_0, \quad y_1 \equiv y_0, \quad F_{t_0}(x_1, y_1) = t_0 N(\alpha x_1 + \beta y_1) + \frac{1}{t_0} N(\gamma x_1 + \delta y_1) \leqslant 1,$$

and so (A) follows by the theorem of the arithmetic and geometric means. For the solution so obtained,

$$N(\alpha x_1 + \beta y_1) \leqslant \frac{1}{t_0},$$

where $\frac{1}{t_0}$ can be made arbitrarily small for irrational $\alpha^{-1}\beta$.

The equality sign holds in (A), if $\alpha = \delta = 1$, $\beta = \gamma = 0$, and both x_0 and y_0 are singular vertices. Theorem 17 is therefore the best possible result.

Mathematical Department,
University of Manchester.

AN INEQUALITY FOR SUMS OF INDEPENDENT RANDOM VARIABLES

By A. C. OFFORD.

[Received 27 August, 1942.—Read 10 December, 1942.]

1. This note deals with a certain inequality for the sum

$$S = \sum_1^n X_\nu \tag{1.1}$$

of a number of independent random variables. Broadly speaking, we show that the probability of this sum lying between narrow assigned bounds is in general small.

Little seems to be known in this direction. There is of course the well-known inequality of Tchebysheff which asserts that

$$\Pr\left\{|S| \geqslant \lambda \sqrt{\left(\sum_1^n a_\nu^2\right)}\right\} \leqslant \frac{1}{\lambda^2},$$

where† $a_\nu^2 = \mathfrak{M}(X_\nu^2)$. In other words, the sum (1.1) can only rarely be a large multiple of $\left(\sum_1^n a_\nu^2\right)^{\frac{1}{2}}$. The type of inequality that we seek is, however, rather different. We wish to show that

$$\Pr\,(a-x \leqslant S \leqslant a+x) \leqslant \phi(x),$$

where $\phi(x)$ is independent of a and small with x. The function $\phi(x)$ will of course depend on the mean values of the random variables X_ν and the possibility of $\phi(x)$ being small will depend on these mean values.

Littlewood and Offord‡ have already given an inequality of this type

† Here $\mathfrak{M}(X^p)$ denotes the mean value of X^p; cf. Cramér, "Random variables and probability distributions", *Cambridge Tracts in Math.*, No. 36 (1937), 19. Cramér uses $E(X^p)$ in place of our $\mathfrak{M}(X^p)$.

‡ Littlewood and Offord, "On the number of real roots of a random algebraic equation III", *Recueil Math. N. S.*, 12 (1943), 277–286.

for Bernoulli convolutions, that is for sums of the form $\sum_1^n \epsilon_\nu a_\nu$, where the a_ν's are given numbers and ϵ_ν is a random variable which takes the values $+1$ and -1 with equal probability. They point out that even in the Bernoulli case $\phi(x)$ cannot be a simple multiple of $\left\{\sum_1^n a_\nu^2\right\}^{\frac{1}{2}}$, and this is all the more true in the general case considered here. If, for instance,

$$X_\nu = \begin{cases} \nu+1 & \text{with probability } \frac{1}{2}(\nu+1)^{-2}, \\ 0 & \text{,,} \quad \text{,,} \quad 1-(\nu+1)^{-2}, \\ -(\nu+1) & \text{,,} \quad \text{,,} \quad \frac{1}{2}(\nu+1)^{-2}, \end{cases} \tag{1.2}$$

then plainly $\mathfrak{M}(X_\nu^2) = 1$, while it is easy to see that

$$\Pr(|S| = 0) \geqslant \prod_{\nu=2}^{\infty} \left(1 - \frac{1}{\nu^2}\right) = \tfrac{1}{2}$$

for large n.

We first introduce the following function which greatly facilitates both the statements and the proofs of our theorems.

DEFINITION†. *If $F(u)$ is any distribution function‡, we define an associated function $\psi(x)$ by the relation*

$$\psi(x) = \max_{-\infty < t < \infty} \int_{t-x}^{t+x} dF(u). \tag{1.3}$$

So, if X is any random variable whose distribution function is $F(u)$, then

$$\Pr(t-x \leqslant X \leqslant t+x) \leqslant \psi(x)$$

and it will therefore be sufficient for us to find an inequality for the ψ function associated with the sum (1.1).

Our principal result is

THEOREM 1. *Suppose that the random variables $X_1, X_2, \ldots, X_n$ are mutually independent, and let*

$$a_\nu = \mathfrak{M}(X_\nu), \quad A_\nu^2 = \mathfrak{M}\{(X_\nu - a_\nu)^2\}, \quad B_\nu^3 = \mathfrak{M}\{|X_\nu - a_\nu|^3\}. \tag{1.4}$$

† This function was first introduced by Littlewood and Offord, *loc. cit.* Lemma 2 and a special case of Lemma 6 were given by them.

‡ Cf. Cramér, *loc. cit.*, p. 11.

Suppose further that

$$2\kappa^3 = \min_\nu \frac{A_\nu}{B_\nu}, \tag{1.5}$$

and let $\psi_n(x)$ *be the* ψ *function associated with the sum* $\sum_1^n X_\nu$. *Then for all* $n > 1$

$$\psi_n(x) \leqslant C \frac{\log n}{\kappa^3 \sqrt{n}} \left\{ \log n + \frac{\kappa x}{\min_\nu A_\nu} \right\}, \tag{1.6}$$

where C *is an absolute constant.*

It may be remarked that by Hölder's inequality $A_\nu \leqslant B_\nu$, and so $\kappa \leqslant \frac{1}{3}$. Also since $\psi_n(x) \leqslant 1$ for all n and x, the inequality (1.6) is non-trivial only when n is large. The theorem is true with $C = 6$.

Theorem 1 is open to an obvious objection. The second member of (1.6) ceases to be small when one of the random variables has either a large third moment or a small second moment. This objection is removed by the following theorem which, in reality, is only a corollary of Theorem 1.

THEOREM 2. *Let the random variables* $X_1, X_2, \ldots, X_n$ *be mutually independent and let* a_ν, A_ν *and* B_ν *be defined as in* (1.4). *Let* s *denote any subsequence* $\nu_1, \nu_2, \ldots, \nu_m$ *out of the* $2^n - n$ *possible subsequences of at least two terms which can be formed from the natural numbers* 1, 2, 3, ..., n. *Write*

$$2\kappa_s^3 = \min_{\nu = \nu_1, \nu_2, \ldots, \nu_m} \frac{A_\nu}{B_\nu}, \quad A^{(s)} = \min_{\nu = \nu_1, \nu_2, \ldots, \nu_m} A_\nu,$$

and let $\psi_n(x)$ *be the* ψ *function of* $\sum_1^n X_\nu$. *Then*

$$\psi_n(x) \leqslant C \min_s \frac{\log m_s}{\kappa_s^3 \sqrt{m_s}} \left\{ \log m_s + \frac{\kappa_s x}{A^{(s)}} \right\},$$

where C *is an absolute constant,* m_s *is the number of terms in the subsequence* s, *and the minimum is taken for all the* $2^n - n$ *possible subsequences* s.

2. The properties of the functions ψ are given by the following lemmas.

LEMMA 1. *For all* $k \geqslant 0$

$$\psi(kx) \leqslant (k+1)\psi(x).$$

This is obvious.

LEMMA 2. *If X and Y are independent random variables, and if ψ_1, ψ_2 and ψ_3 are the ψ functions associated with X, Y and $X+Y$ respectively, then*

$$\psi_3(x) \leqslant \min\{\psi_1(x),\ \psi_2(x)\}.$$

Proof. Let $F_1(u)$, $F_2(u)$ and $F_3(u)$ be the distribution functions of X, Y, and $X+Y$ respectively. Then

$$F_3(t) = \int_{-\infty}^{\infty} F_2(t-u)\, dF_1(u).$$

Hence

$$|F_3(t+x)-F_3(t-x)| \leqslant \int_{-\infty}^{\infty} |F_2(t-u+x)-F_2(t-u-x|\, dF_1(u)$$

$$\leqslant \psi_2(x) \int_{-\infty}^{\infty} dF_1(u) = \psi_2(x).$$

Therefore $$\psi_3(x) \leqslant \psi_2(x),$$

and similarly $\psi_3(x) \leqslant \psi_1(x)$ as desired.

The inequality established in the following lemma may seem at first sight a little odd, but the reader must bear in mind that, in this case, $A < B < \infty$ and $\kappa \leqslant \frac{1}{8}$, as remarked after Theorem 1.

LEMMA 3†. *If $\mathfrak{M}(X) = a$, $\mathfrak{M}\{(X-a)^2\} = A^2$, $\mathfrak{M}\{|X-a|^3\} = B^3$, where B^3 is finite, and if $A \geqslant 2\kappa^{\frac{1}{3}} B$, then, $\psi(x)$ being the ψ function associated with X,*

$$\psi(\kappa A) \leqslant 1-2\kappa^2.$$

Proof. Since the ψ functions corresponding to X and $X-a$ are identical, we may without loss of generality suppose that $a = 0$. Then, if $F(u)$ is the distribution function of X, we have

$$a = \int_{-\infty}^{\infty} u\, dF(u) = 0,$$

so that, if

$$\beta = \mathfrak{M}(|X|) = \int_{-\infty}^{\infty} |u|\, dF(u),$$

we get

$$\int_0^{\infty} u\, dF(u) = \tfrac{1}{2}\beta. \tag{2.1}$$

† The proof of this lemma uses the idea of a theorem of Zygmund, *Mathesis Polska*, 8 (1933), 76–87.

Now, by Schwarz's inequality,

$$\mathfrak{M}(X^2) \leqslant \{\mathfrak{M}(|X|^3)\,\mathfrak{M}(|X|)\}^{\frac{1}{2}} = B^{\frac{3}{2}}\beta^{\frac{1}{2}},$$

so that

$$\beta \geqslant A^4 B^{-3} \geqslant 8\kappa A, \tag{2.2}$$

in virtue of the hypothesis of the lemma.

Again, for any $\Delta > 0$, we have

$$\int_{\Delta}^{\infty} u\,dF(u) \leqslant \frac{1}{\Delta}\int_{\Delta}^{\infty} u^2\,dF(u) \leqslant \frac{A^2}{\Delta}. \tag{2.3}$$

Also

$$\int_0^{\delta} u\,dF(u) \leqslant \delta\int_0^{\infty} dF(u) \leqslant \delta. \tag{2.4}$$

Therefore, combining (2.1), (2.2), (2.3), and (2.4), we deduce

$$\begin{aligned}\int_{\kappa A}^{A/\kappa} u\,dF(u) &= \int_0^{\infty} u\,dF(u) - \int_{A/\kappa}^{\infty} u\,dF(u) - \int_0^{\kappa A} u\,dF(u)\\ &\geqslant 4\kappa A - \kappa A - \kappa A = 2\kappa A,\end{aligned}$$

and this implies

$$\int_{\kappa A}^{\infty} dF(u) \geqslant \frac{\kappa}{A}\int_{\kappa A}^{A/\kappa} u\,dF(u) \geqslant 2\kappa^2.$$

Hence, for $t < 0$,

$$\int_{t-\kappa A}^{t+\kappa A} dF(u) \leqslant \int_{-\infty}^{\kappa A} dF(u) \leqslant 1 - 2\kappa^2,$$

and similarly for $t > 0$.

LEMMA 4. *Suppose that X and Y are independent random variables and that $\mathfrak{M}(Y) = \alpha$, $\mathfrak{M}\{(Y-\alpha)^2\} = A^2$. Further, let $\psi_1(x)$, $\psi_2(x)$ and $\psi_3(x)$ be the ψ functions of X, Y and $X+Y$ respectively; then*

$$\psi_3(x) \leqslant \psi_2(x)\,\psi_1(\lambda A) + \frac{1}{\{\lambda - (x/A)\}^2},$$

provided that $\lambda > 1$ and $x/A < \lambda$.

Proof. As in Lemma 3, without loss of generality we may suppose that $\alpha = 0$. Let $F_1(u)$, $F_2(u)$ and $F_3(u)$ be the distribution functions of X, Y

and $X+Y$. Then

$$F_3(x) = \int_{-\infty}^{\infty} F_2(x-u)\, dF_1(u),$$

and so

$$\begin{aligned} F_3(t+x) - F_3(t-x) &= \int_{-\infty}^{\infty} \{F_2(t-u+x) - F_2(t-u-x)\}\, dF_1(u) \\ &= \int_{-\infty}^{\infty} \{F_2(-y+x) - F_2(-y-x)\}\, dF_1(y+t) \\ &= \int_{-\lambda A}^{\lambda A} + \int_{\lambda A}^{\infty} + \int_{-\infty}^{-\lambda A} \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Now

$$I_1 \leqslant \psi_2(x) \int_{t-\lambda A}^{t+\lambda A} dF_1(y) \leqslant \psi_2(x)\, \psi_1(\lambda A). \tag{2.5}$$

Again, since by hypothesis $y \geqslant \lambda A > x$,

$$\begin{aligned} F_2(-y+x) - F_2(-y-x) &= \int_{-y-x}^{-y+x} dF_2(u) \leqslant \int_{-\infty}^{-y+x} dF_2(u) \\ &\leqslant \frac{1}{(y-x)^2} \int_{-\infty}^{-y+x} u^2\, dF_2(u) \leqslant \frac{A^2}{(y-x)^2}. \end{aligned}$$

Therefore

$$I_2 \leqslant \int_{\lambda A}^{\infty} \frac{1}{(y-x)^2}\, dF_1(y+t) \leqslant \left(\lambda - \frac{x}{A}\right)^{-2} \int_{\lambda A}^{\infty} dF_1(y+t).$$

Similarly

$$I_3 \leqslant \left(\lambda - \frac{x}{A}\right)^{-2} \int_{-\infty}^{-\lambda A} dF_1(y+t),$$

and so

$$I_2 + I_3 \leqslant \left(\lambda - \frac{x}{A}\right)^{-2}. \tag{2.6}$$

Combining (2.5) and (2.6), we see that

$$\begin{aligned} \psi_3(x) &= \max_t \{F_3(t+x) - F_3(t-x)\} \\ &\leqslant \psi_2(x)\, \psi_1(\lambda A) + \left(\lambda - \frac{x}{A}\right)^2, \end{aligned}$$

which is the desired result.

LEMMA 5. *Suppose that $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables, and let a_ν, A_ν and B_ν be defined by* (1.4). *Suppose further that*

$$A_\nu \geqslant 2\kappa^{\frac{1}{2}} B_\nu \tag{2.7}$$

for all ν, and

$$\frac{A_{\nu+1}}{A_\nu} \leqslant \frac{\kappa^2(1-\kappa^2)^{\frac{1}{2}\nu}}{1+\kappa^2} \tag{2.8}$$

for all ν. Then, if $\psi_n(x)$ is the ψ function associated with the sum $\sum_1^n X_\nu$, we have

$$\psi_n(\kappa A_n) \leqslant (1-\kappa^2)^n$$

for all n.

Proof. The proof is by induction. The result is true for $n=1$ in virtue of Lemma 3. We suppose it true for $n=m$, and show that it will then follow for $n=m+1$. Let $\psi^*_{m+1}(x)$ be the ψ function of X_{m+1}. In Lemma 4 take $A=A_{m+1}$ and $x=\kappa A_{m+1}$. Then

$$\begin{aligned}\psi_{m+1}(\kappa A_{m+1}) &\leqslant \psi^*_{m+1}(\kappa A_{m+1})\,\psi_m(\lambda A_{m+1}) + (\lambda-\kappa)^{-2} \\ &\leqslant (1-2\kappa^2)\,\psi_m(\lambda A_{m+1}) + (\lambda-\kappa)^{-2},\end{aligned}$$

in virtue of Lemma 3.

Now let

$$\lambda = \lambda_m = \frac{1+\kappa^2(1-\kappa^2)^{\frac{1}{2}m}}{\kappa(1-\kappa^2)^{\frac{1}{2}m}}.$$

Then, using (2.8), we have

$$\lambda_m A_{m+1} < \frac{1+\kappa^2}{\kappa(1-\kappa^2)^{\frac{1}{2}m}} A_{m+1} \leqslant \kappa A_m.$$

Also

$$\lambda_m - \kappa = \frac{1}{\kappa(1-\kappa^2)^{\frac{1}{2}m}}.$$

Hence, using the hypothesis of induction, we have

$$\begin{aligned}\psi_{m+1}(\kappa A_{m+1}) &\leqslant (1-2\kappa^2)\,\psi_m(\kappa A_m) + \kappa^2(1-\kappa^2)^m \\ &\leqslant (1-2\kappa^2)(1-\kappa^2)^m + \kappa^2(1-\kappa^2)^m \\ &= (1-\kappa^2)^{m+1},\end{aligned}$$

and this proves the lemma.

LEMMA 6. *Suppose that $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables and let a_ν, A_ν and B_ν be defined by* (1.4). *Suppose further*

that (2.7) *is satisfied, and in addition that*

$$\tfrac{1}{2}K \leqslant A_\nu \leqslant K \tag{2.9}$$

for all ν. *Then, if* $\psi_n(x)$ *is the* ψ *function of* $\sum_1^n X_\nu$,

$$\psi_n(x) \leqslant \frac{6}{\sqrt{n}}\left\{\frac{\log n}{\kappa} + \sqrt{\left(\frac{2}{\pi}\right)}\,\frac{x}{K}\right\}.$$

Proof. As in Lemma 3, without loss of generality we may suppose that $a_\nu = 0$ for all ν. Let $\sigma_n = \left\{\sum_1^n A_\nu^2\right\}^{\frac{1}{2}}$, and let $F_n(u)$ be the distribution function of $\frac{1}{\sigma_n}\sum_1^n X_\nu$. Then, by Liapounoff's theorem†,

$$\left|F_n(t) - \frac{1}{\sqrt{(2\pi)}\,\sigma_n}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2/\sigma_n^2}\,dt\right| \leqslant 3\,\frac{\sum_1^n B_\nu^3}{\left(\sum_1^n A_\nu^2\right)^{\frac{3}{2}}}\log n \leqslant \frac{3}{8\kappa}\,\frac{\sum_1^n A_\nu^3}{\left(\sum_1^n A_\nu^2\right)^{\frac{3}{2}}}\log n$$

$$\leqslant \frac{3}{8\kappa}\,\frac{K^3}{(\frac{1}{2}K)^3}\,\frac{\log n}{\sqrt{n}} \leqslant \frac{3\log n}{\kappa\sqrt{n}}.$$

Hence

$$\psi_n(x) = \max_y\{F_n(y+x) - F_n(y-x)\}$$

$$\leqslant \frac{6\log n}{\kappa\sqrt{n}} + \frac{1}{\sqrt{(2\pi)}\,\sigma_n}\max_y\int_{y-x}^{y+x} e^{-\frac{1}{2}u^2/\sigma_n^2}\,du$$

$$\leqslant \frac{6\log n}{\kappa\sqrt{n}} + \frac{2}{\sqrt{(2\pi)}}\,\frac{x}{\sigma_n} \leqslant \frac{6}{\sqrt{n}}\left\{\frac{\log n}{\kappa} + \frac{2}{3\sqrt{(2\pi)}}\,\frac{x}{K}\right\}.$$

3. *Proof of Theorem* 1. Arrange the A_ν's in order of decreasing magnitude and divide them into p^2 blocks of at least $[n/p^2]$ terms each. Let A_n and $A_{n_i'}$ be the first and last terms of a typical block. Either

$$A_{n_i'} < \tfrac{1}{2}A_{n_i} \tag{3.1}$$

for all i, or else there is a block, say the j-th, for which

$$A_{n_j'} \geqslant \tfrac{1}{2}A_{n_j}. \tag{3.2}$$

† Cf. Cramér, *loc. cit.*, p. 77.

Let us take case (3.2) first. With the X_ν's arranged according to decreasing A_ν, write

$$\sum_1^n X_\nu = \sum_{\nu=n_j}^{n_j'} X_\nu + \Sigma' = \Sigma_1 + \Sigma_2,$$

and let ψ_n and ψ^* be the ψ functions corresponding to $\sum_1^n X_\nu$ and Σ_1 respectively. Writing m for $n_j' - n_j + 1$ we have, in virtue of Lemmas 2 and 6,

$$\psi_n(x) \leqslant \psi^*(x) \leqslant \frac{6}{\sqrt{m}} \left\{ \frac{\log m}{\kappa} + \sqrt{\left(\frac{2}{\pi}\right)} \frac{x}{3A_{n_j}} \right\}. \tag{3.3}$$

Now consider the case (3.1). We first make a further selection from the A's. Let h_j be an integer such that

$$2^{h_j - 1} < \frac{1+\kappa^2}{\kappa^2(1-\kappa^2)^{\frac{1}{2}j}} \leqslant 2^{h_j}. \tag{3.4}$$

We then define the sequence

$$m_1 = n_1, \quad m_2 = n_{h_1}, \quad m_3 = n_{h_1+h_2}, \quad \ldots, \quad m_j = n_{h_1+h_2\ldots+h_{j-1}}, \quad \ldots.$$

Then

$$A_{m_{j+1}} \leqslant 2^{-h_j} A_{m_j} \leqslant \frac{\kappa^2(1-\kappa^2)^{\frac{1}{2}j}}{1+\kappa^2} A_{m_j}.$$

Let m_q be the last m_j that can be chosen from the sequence

$$n_1, \ n_2, \ \ldots, \ n_{p^2}.$$

Then

$$\sum_{j=1}^{q} h_j > p^2. \tag{3.5}$$

We now establish an inequality for q. Since $\kappa \leqslant \frac{1}{8}$, by (3.4), we have

$$\begin{aligned}(h_j - 1)\log 2 &\leqslant \log \frac{1+\kappa^2}{\kappa^2(1-\kappa^2)^{\frac{1}{2}j}} \\ &\leqslant 2\log\frac{1}{\kappa} + \log\tfrac{65}{64} + \tfrac{1}{2}j\log\tfrac{64}{63},\end{aligned}$$

i.e.,

$$h_j \log 2 \leqslant 2\log\frac{1}{\kappa} + \frac{j}{126} + \log\tfrac{65}{32}.$$

Hence, by (3.5),

$$p^2 \log 2 \leqslant 2q \log \frac{1}{\kappa} + \frac{q(q+1)}{126} + q \log \tfrac{6}{3}\tfrac{5}{2}.$$

Therefore

$$p^2 \leqslant \tfrac{7}{2} q \log \frac{1}{\kappa} + \tfrac{1}{8} q^2,$$

and so, if $\kappa \geqslant e^{-\frac{1}{4}p}$, we must have

$$p^2 \leqslant \tfrac{7}{8} qp + \tfrac{1}{8} q^2,$$

which implies $q \geqslant p$.

Hence, if $\psi^{**}(x)$ denotes the ψ function corresponding to the sum $\sum_{j=1}^{q} X_{m_j}$, we have, by Lemma 5,

$$\psi^{**}(x) \leqslant (1-\kappa^2)^q \leqslant (1-\kappa^2)^p,$$

provided that $x \leqslant \kappa A_{m_j}$, and so, by Lemma 2,

$$\psi_n(\kappa A_n) \leqslant \psi^{**}(\kappa A_n) \leqslant (1-\kappa^2)^p,$$

provided that $\kappa \geqslant e^{-\frac{1}{4}p}$. Hence, if $x > \kappa A_n$, we have, by Lemma 1,

$$\psi_n(x) \leqslant \left(1 + \frac{x}{\kappa A_n}\right)(1-\kappa^2)^p. \tag{3.6}$$

Since the A_ν's are arranged in decreasing order, we have also $A_{n_j} \geqslant A_n$. Therefore, combining (3.3) and (3.6), we have, in either case,

$$\psi_n(x) = \max\left[\frac{6}{\sqrt{m}}\left\{\frac{\log m}{\kappa} + \sqrt{\left(\frac{2}{\pi}\right)\frac{x}{3A_n}}\right\},\ \left(1+\frac{x}{\kappa A_n}\right)(1-\kappa^2)^p\right],$$

where $m = [n/p^2]$.

Taking $p = [\kappa^{-2} \log n]$, so that the condition $\kappa \geqslant e^{-\frac{1}{4}p}$ is certainly satisfied, we get

$$\begin{aligned}
\psi_n(x) &\leqslant \max\left[\left\{\frac{6\log^2 n}{\kappa^3\sqrt{n}} + \sqrt{\left(\frac{2}{\pi}\right)\frac{2x\log n}{A_n\kappa^2\sqrt{n}}}\right\};\ \left\{\left(1+\frac{x}{\kappa A_n}\right)\frac{1}{n}\right\}\right] \\
&\leqslant \frac{6\log^2 n}{\kappa^3\sqrt{n}} + \sqrt{\left(\frac{2}{\pi}\right)}\frac{2\log n}{\kappa^2\sqrt{n}}\frac{x}{A_n} \\
&\leqslant \frac{6\log n}{\kappa^3\sqrt{n}}\left(\log n + \frac{\kappa x}{A_n}\right),
\end{aligned}$$

and this proves the theorem.

Theorem 2 is, in view of Lemma 2, only a corollary of Theorem 1. For if $\nu_1, \nu_2, \ldots, \nu_m$ is any subsequence of the natural numbers $1, 2, \ldots, n$, then we may write

$$S = \sum_{i=1}^{m} X_{\nu_i} + \Sigma' = \Sigma_1 + \Sigma_2,$$

and if $\psi^*(x)$ is the ψ function of Σ_1, we have, by Lemma 2,

$$\psi_n(x) \leqslant \psi^*(x).$$

But, by Theorem 1,

$$\psi^*(x) \leqslant C \frac{\log m}{\kappa_s^3 \sqrt{m}} \left(\log m + \frac{\kappa_s x}{A_s}\right),$$

and since this holds for all subsequences $\{\nu_i\}$ the theorem follows.

King's College,
Newcastle-on-Tyne.

PROCEEDINGS

OF

THE LONDON MATHEMATICAL SOCIETY.

SERIES 2.—VOL. 48.—PART 1.

CONTENTS.

THE BUCKLING OF GRIDS OF STRINGERS AND RIBS: H. L. COX *and* H. E. SMITH 1

THE BUCKLING OF A THIN SHEET TRANSVERSELY STIFFENED: H. L. COX *and* H. E. SMITH 27

A UNIQUENESS THEOREM FOR MEAN VALUES OF ANALYTIC FUNCTIONS: W. H. J. FUCHS 35

ON THE COEFFICIENTS OF SUBORDINATE FUNCTIONS: WERNER ROGOSINSKI 48

Issued June 25th, 1943

LONDON:

PRINTED AND PUBLISHED FOR THE SOCIETY BY

C. F. HODGSON & SON, LTD.,

2 NEWTON STREET, KINGSWAY, W.C.2.

PRICES OF
The London Mathematical Society's Publications.

(Revised May 16th, 1932.)

PROCEEDINGS (First Series).—The first series consists of 35 volumes published between January, 1865, and January, 1903. The price per volume (cloth, 8vo) ranges from 17*s.* 6*d.* to 45*s.* according to the size of the volume and the number of copies in stock (postage 1*s.* per volume extra); prices of individual volumes may be ascertained by inquiry of the Publishers. The price of the set of 35 volumes (not including carriage) is £56. 14*s.*, subject to the discount stated below.

[NOTE.—Vols. I and II of the first series will not be sold except to purchasers of not less than twenty volumes of that series.]

PROCEEDINGS (Second Series).—Vol. I of the second series was published in 1903; 47 volumes have been issued to date (1942); most volumes up to XLIII contain about 560 pages; subsequent volumes will contain about 480 pages. The price per volume (cloth, Super Royal 8vo) is 35*s.* (postage 1*s.* per volume extra).

INDEXES.—*Complete Index* (112 pages) to Vols. I–XXX of Series 1 of the *Proceedings*, 5*s.* (by post, 5*s.* 3*d.*). *Complete Index* (143 pages) to Vols. XXXI–XXXV of Series 1 and Vols. I–XXX of Series 2 of the *Proceedings*, 5*s.* (by post, 5*s.* 3*d.*).

JOURNAL.—Vol. I was published in 1926; sixteen volumes have been issued to date (1942); recent volumes contain about 320 pages. The price per volume (cloth, Super Royal 8vo) is 20*s.* (postage 1*s.* per volume extra).

SUBSCRIPTION PRICES.—The public may subscribe to the current volume of the *Proceedings* (price 35*s.*) and to the current volume of the *Journal* (price 20*s.*), these sums being payable in advance, and can receive them either in PARTS (about six to volume of *Proceedings* and four to volume of *Journal*) AS ISSUED, or, when ready, as BOUND VOLUMES, post free.

SEPARATE PARTS.—Separate parts of both *Proceedings* and *Journal* may be purchased at 7*s.* 6*d.* each (by post, 8*s.*).

BINDING CASES.—Cloth binding cases for either series of the *Proceedings* and for the *Journal* may be purchased at 2*s.* each (by post 2*s.* 3*d.*).

LIST OF MEMBERS (current issue only), 5*s.* each (by post 5*s.* 3*d.*).

NOTES ON THE PREPARATION OF MATHEMATICAL PAPERS (19 pages), 1*s.*, post free. If 50 or more copies are purchased, the price is at the rate of 30*s.* per 100, post free.

DISCOUNT.—A discount of 10 per cent. off the above prices is allowed (i) to Members of the Society purchasing separate parts or volumes to complete their own sets, (ii) to any person purchasing a run of 20 or more volumes of the *Proceedings*. Both discounts are allowed when both are applicable. With these exceptions, the prices quoted above are net.

REPRINTS OF AUTHORS' PAPERS.—Authors of Papers printed in the *Proceedings* or *Journal* are entitled to receive 25 separate copies of their papers free of cost. They can, however, by previous notice to the Printers, obtain not more than 50 copies at the approximate cost of production. *If* **any** *separate copies are required, an Order Form must be filled in and returned to the Printers with proofs.*

C. F. HODGSON & SON, LTD. (Publishers to the London Mathematical Society), 2 NEWTON STREET, W.C.2
Telephone: Hol. 1305–1306. Bankers: National Provincial.